OPTIMAL TRANSPORT AND SKOROKHOD EMBEDDING

MATHIAS BEIGLBÖCK, ALEXANDER M. G. COX, AND MARTIN HUESMANN

ABSTRACT. The *Skorokhod embedding problem* is to represent a given probability as the distribution of Brownian motion at a chosen stopping time. Over the last 50 years this has become one of the important classical problems in probability theory and a number of authors have constructed solutions with particular optimality properties. These constructions employ a variety of techniques ranging from excursion theory to potential and PDE theory and have been used in many different branches of pure and applied probability.

We develop a new approach to Skorokhod embedding based on ideas and concepts from *optimal mass transport*. In analogy to the celebrated article of Gangbo and McCann on the geometry of optimal transport, we establish a geometric characterization of Skorokhod embeddings with desired optimality properties. This leads to a systematic method to construct optimal embeddings. It allows us, for the first time, to derive all known optimal Skorokhod embeddings as special cases of one unified construction and leads to a variety of new embeddings. While previous constructions typically used particular properties of Brownian motion, our approach applies to all sufficiently regular Markov processes.

Keywords: Optimal Transport, Skorokhod Embedding, cyclical monotonicity.
Mathematics Subject Classification (2010): Primary 60G42, 60G44; Secondary 91G20.

1. INTRODUCTION

Let B be a Brownian motion started in 0 and consider a probability μ on the real line which is centered and has second moment. The Skorokhod embedding problem is to construct a stopping time τ embedding μ into Brownian motion in the sense that

$$B_\tau \text{ is distributed according to } \mu, \quad \mathbb{E}[\tau] < \infty. \quad \text{(SEP)}$$

Here, the second condition is imposed to exclude certain undesirable solutions. It is not hard to see that $\mathbb{E}[\tau] = \int x^2 \mu(dx)$ for any solution of (SEP). As already demonstrated by Skorokhod [53, 54] in the early 1960's, it is always possible to construct solutions to the problem. Indeed, the survey article of Obłój classifies 21 distinct solutions to (SEP), although this list (from 2004) misses many more recent contributions. A common inspiration for many of these papers is to construct solutions to (SEP) that exhibit additional desirable properties or a distinct internal structure. These have found applications in different fields and various extensions of the original problem have been considered. We refer to the survey of Obłój [39] (and the 120+ references therein) for a comprehensive account of the field.

Our aim is to develop a new approach to (SEP) based on ideas from optimal transport. Many of the previous developments are thus obtained as applications of one unifying principle (Theorem 1.2) and several difficult problems are rendered tractable.

1.1. A motivating example — Root's construction.
To illustrate our approach we introduce a solution that will serve as inspiration in the rest of the paper: Root's construction [46] which is one of the earliest solutions to (SEP). It is prototypical for many further solutions to (SEP) in that it has a simple *geometric description* and possesses a certain *optimality property* in the class of all solutions.
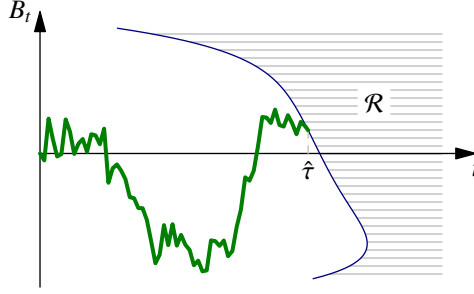
FIGURE 1. Root's solution of (SEP).

Root established that there exists a *barrier* $\mathcal{R}$ (which is essentially unique) such that the Skorokhod embedding problem is solved by the stopping time

$$\tau_{\text{Root}} = \inf\{t \geq 0 : (t, B_t) \in \mathcal{R}\}. \tag{1.1}$$

A barrier is a Borel set $\mathcal{R} \subseteq \mathbb{R}_+ \times \mathbb{R}$ such that $(s, x) \in \mathcal{R}$ and $s < t$ implies $(t, x) \in \mathcal{R}$. The Root construction is distinguished by the following optimality property: among all solutions to (SEP) for a fixed terminal distribution μ, it minimizes $\mathbb{E}[\tau^2]$. For us, the optimality property will be the starting point from which we deduce a geometric characterization of τ_{Root}. To this end, we now formalize the corresponding optimization problem.

1.2. **Optimal Skorokhod Embedding Problem.** We consider the set of *stopped paths*

$$S = \{(f, s) : f : [0, s] \to \mathbb{R} \text{ is continuous}, f(0) = 0\}. \tag{1.2}$$

Throughout the paper we consider a functional

$$\gamma : S \to \mathbb{R}.$$

The optimal Skorokhod embedding problem is to construct a stopping time optimizing

$$P_\gamma(\mu) = \inf\left\{\mathbb{E}[\gamma((B_t)_{t\leq\tau}, \tau)] : \tau \text{ solves (SEP)}\right\}. \tag{OptSEP}$$

We will usually assume that (OptSEP) is *well-posed* in the sense that $\mathbb{E}\left[\gamma((B_t)_{t\leq\tau}, \tau)\right]$ exists with values in $(-\infty, \infty]$ for all τ which solve (SEP) and is finite for one such τ.

The Root stopping time solves (OptSEP) in the case where $\gamma(f, s) = s^2$. Other examples where the solution is known include functionals depending on the running maximum $\gamma((f, s)) := \bar{f}(s) := \max_{t\leq s} f(t)$ or functionals of the local time at 0.

The solutions to (SEP) have their origins in many different branches of probability theory, and in many cases, the original derivation of the embedding occurred separately from the proof of the corresponding optimality properties. Moreover, the optimality of a given construction is often not entirely obvious; for example, the optimality property of the Root embedding was first conjectured by Kiefer [30] and subsequently established by Rost [48].

In contrast to existing work, our starting point will be the optimization problem (OptSEP) and we seek a systematic method to construct the minimizer for a given functional γ. To develop a general theory for this optimization problem we interpret stopping times in terms of a transport plan from the Wiener space $(C(\mathbb{R}_+), \mathbb{W})$ to the target measure μ, i.e. we want to think of a stopping time τ as transporting the mass of a trajectory $(B_t(\omega))_{t\in\mathbb{R}_+}$ to the point $B_{\tau(\omega)}(\omega) \in \mathbb{R}$. Note that this is *not a coupling* between $\mathbb{W}$ and μ in the usual sense and one cannot directly apply optimal transport theory. Instead we develop an analogous theory, which in particular needs to account for the adaptedness properties of stopping times. To this end, it is necessary to combine ideas and results from optimal transportation with concepts and techniques from stochastic analysis.

As in optimal transport, it is crucial to consider (OptSEP) in a suitably relaxed form, i.e. in (OptSEP) we will optimize over *randomized stopping times* (see Theorem 4.16 below). These can be viewed as usual stopping times on a possibly enlarged probability space but in our context it is more natural to interpret them as "Kantorovich-type" stopping times,

i.e. stopping times which terminate a given path not at a single deterministic time instance but according to a distribution.

This relaxation will allow us to transfer many of the convenient properties of classical transport theory to our probabilistic setup. Exactly as in classical transport theory, (OptSEP) can be viewed as a linear optimization problem. The set of couplings in mass transport is compact and similarly the set of all randomized stopping times solving (SEP) is compact in a natural sense. As a particular consequence we will establish:

Theorem 1.1. *Let $\gamma : S \rightarrow \mathbb{R}$ be lower semi-continuous and bounded from below. Then* (OptSEP) *admits a minimizer $\hat{\tau}$. More precisely, there exists a Brownian motion B on some stochastic basis $\Omega = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and a stopping time $\hat{\tau}$ of B which attains* (OptSEP).

Here we can talk about the continuity properties of γ since S possesses a natural Polish topology (cf. (4.1)).

In terms of linear optimization, Theorem 1.1 is a primal problem. In Section 5 we will introduce the corresponding dual problem and establish that there is no duality gap.

1.3. **Geometric Characterization of Optimizers — Montonicity Principle.** A fundamental idea in optimal transport is that the optimality of a transport plan is reflected by the geometry of its support set. Often this is key to understanding the transport problem. On the level of support sets, the relevant notion is *c-cyclical monotonicity* which we recall in (3.4) below. Its relevance for the theory of optimal transport has been fully recognized by Gangbo and McCann [21], based on earlier work of Knott and Smith [32] and Rüschendorf [49, 50] among others.

Inspired by these results, we establish a *monotonicity principle* which links the optimality of a stopping time τ with "geometric" properties of τ. Combined with Theorem 1.1, this principle will turn out to be surprisingly powerful. For the first time, *all* the known solutions to (SEP) with optimality properties can be established through one unifying principle. Moreover, the monotonicity principle allows us to treat the optimization problem (OptSEP) in a systematic manner, generating further embeddings as a by-product.

Importantly, this transport-based approach readily admits a number of strong generalizations and extensions. With only minor changes our existence result, Theorem 1.1, and the monotonicity principle, Theorem 1.2, extend to general starting distributions and Brownian motion in $\mathbb{R}^d$ and more generally to sufficiently regular Markov processes; see Sections 6 and 8. This is notable since previous constructions usually exploit rather specific properties of Brownian motion.

Theorem 1.2 (Monotonicity Principle)**.** *Let $\gamma : S \rightarrow \mathbb{R}$ be Borel measurable[1], B be a Brownian motion on some stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and τ an optimizer of* (OptSEP). *Then, there exists a γ-monotone Borel set $\Gamma \subseteq S$ such that $\mathbb{P}$-a.s.*

$$((B_t)_{t \leq \tau}, \tau) \in \Gamma . \tag{1.3}$$

If (1.3) holds, we will loosely say that Γ *supports* τ. The significance of Theorem 1.2 is that it links the optimality of the stopping time τ with a particular property of the set Γ, i.e. γ-monotonicity. In applications, the latter turns out to be much more tangible.

The precise definition of γ-*monotonicity* is intricate and we present it in its simplest and most strict form in this introductory section. (See Definition 6.4 / Section 7.1 for a more general version that leads to a stronger assertion in Theorem 1.2.) To link the optimality of a stopping time with properties of the set Γ we consider the minimization problem (OptSEP) on a pathwise level. Consider two paths $(f, s), (g, t) \in S$ which end at the same level, i.e. $f(s) = g(t)$. We want to determine which of the two paths should be "stopped" and which one should be allowed to "go" on further, bearing in mind that we try to minimize the functional in (OptSEP). Specifically we will call $((f, s), (g, t))$ a *stop-go* pair if it is advantageous to stop (f, s) and to go on after (g, t) in the following strong sense:

---

[1]We emphasize that we do not require continuity assumptions on γ here. This will be important when we apply our results.

Definition 1.3. *We say that* $((f, s), (g, t)) \in S \times S$ *is a* stop-go *pair iff* $f(s) = g(t)$ *and*

$$\gamma((f \oplus h, s + u)) + \gamma((g, t)) \quad > \quad \gamma((f, s)) + \gamma((g \oplus h, t + u)) \qquad (1.4)$$

holds[2] for all $(h, u) \in S$ *with* $u > 0$. *The set of stop-go pairs will be denoted by* SG.

In the case where $\gamma(f, s) = s^2$, (1.4) becomes $(s + u)^2 + t^2 > s^2 + (t + u)^2$ which is true iff $s > t$. Hence, in the case of the Root embedding, a stop-go pair $((f, s), (g, t))$ is characterized by $f(s) = g(t)$, $s > t$ (cf. the right side of Figure 2). The set $\Gamma \subseteq S$ contains all the possible "stopped" paths: that is, a path (g, t) is in Γ if there is some possibility that the optimal stopping rule decides to stop at time t having observed the path $(g(r))_{r \in [0,t]}$. Corresponding to the set of stopped paths is the set of paths which we may observe, and at which we may not yet have stopped: these are the still "going" paths. Since all paths must eventually be stopped, we deduce that a path may be going if there is a longer, stopped path which contains the going path as a sub-path. Specifically, if $(\tilde{f}, \tilde{s}) \in \Gamma$ is a stopped path, then the sub-paths $(f, s) = ((\tilde{f}(r))_{r \in [0,s]}, s)$ are going for all $s < \tilde{s}$. We write $\Gamma^<$ for the set of going paths corresponding to the stopped paths Γ, so:

$$\Gamma^< := \{(f, s) : \exists (\tilde{f}, \tilde{s}) \in \Gamma, s < \tilde{s} \text{ and } f \equiv \tilde{f} \text{ on } [0, s]\} . \qquad (1.5)$$

We can now formally introduce γ-monotonicity.

Definition 1.4. *A set* $\Gamma \subseteq S$ *is called* γ-monotone *iff* $\Gamma^< \times \Gamma$ *contains no stop-go pairs, i.e.*

$$\text{SG} \cap (\Gamma^< \times \Gamma) = \emptyset. \qquad (1.6)$$

By the monotonicity principle, an optimal stopping time is supported by a set Γ such that $\Gamma^< \times \Gamma$ contains no stop-go pair $((f, s), (g, t))$. Intuitively, such a pair gives rise to a possible modification, improving the given stopping rule: as $f(s) = g(t)$, we can imagine stopping the path (f, s) at time s, and allowing (g, t) to go on by transferring all paths which extend (f, s), the "remaining lifetime", onto (g, t), which is now "going". By (1.4) this guarantees an improved value of γ in total, contradicting the optimality of γ. Observe that the condition $f(s) = g(t)$ is necessary to guarantee that a modified stopping rule still embeds the measure μ. A pictorial representation of this process in the case of the Root embedding is given in Figure 2.
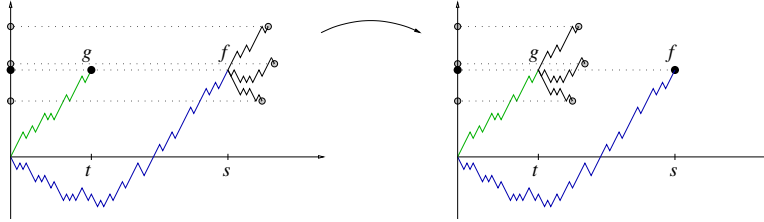

FIGURE 2. How to improve a stopping rule if $\Gamma^< \times \Gamma$ contains a stop-go pair.

In Section 2 below we give a short teaser on how particular embeddings are obtained as a consequence of Theorem 1.2: there we establish the Root and the Rost solutions of (SEP), as well as a continuum of new embeddings which "interpolate" between them. It will become clear that the essence of the proof is already contained in Figure 2. In addition, we are also able to establish many other solutions to the embedding problem, and we proceed to do this in Section 7. For readers who are more interested in the probabilistic consequences of our results, we provide a probabilistic interpretation of our main results in more classical terminology in Section 7.1; this section may be read directly after Section 2.

The monotonicity principle, Theorem 1.2, is the most complex part of this paper, and requires substantial preparation in order to combine the relevant concepts from stochastic

---

[2]Here $f \oplus h$ denotes the concatenation of the two paths f and h. More precisely, $f \oplus h(t) = f(t)$ for $t \in [0, s]$ and $f \oplus h(t) = f(s) + h(t - s)$ for $t \in [s, s + u]$.

analysis and optimal transport. The preparation and proof of this result will therefore comprise the majority of the paper.

The "classical" optimal transport version of Theorem 1.2 can be established through fairly direct arguments, at least in a reasonably regular setting, cf. [3, Thms. 3.2, 3.3] and [57, p. 88f]. However, these approaches do not extend easily to our setup: stopping times are of course not couplings in the usual sense and there is no reason for particular combinatorial manipulations to carry over in a direct fashion. Another substantial difference is that the procedure of transferring paths described below Definition 1.4 necessarily refers to a *continuum* of paths while the classical notion of cyclical monotonicity is concerned with rearrangements along finite cycles. The argument given subsequently is more in the spirit of [6, 8] and requires a fusion of ideas from optimal transport and stochastic analysis. To achieve this, we will need to revisit a number of classical notions from the theory of stochastic processes within a novel framework.

1.4. **New Horizons.** The methods and results presented in this paper are limited to the case of the classical Skorokhod embedding problem for Markov processes with continuous paths. However we believe that our methods are sufficiently general that a number of interesting and important extensions, which previously would have been intractable, may now be within reach:

(1) **Markov processes:** The results presented in this paper should extend to a more general class of Markov processes with càdlàg paths. The main technical issues this would present lie in the generalization of the results in Section 4, where the specific structure of the space of continuous paths is exploited.

(2) **Multiple path-swapping:** In our monotonicity principle, Theorem 1.2, we consider the impact of swapping mass from a single unstopped path onto a single stopped path, and argue that if this improves the objective γ on average, then we cannot observe such behaviour under an optimizer. In classical optimal transport, it is known that single swapping is not sufficient to guarantee optimality; rather, one needs to consider the impact of allowing a finite "cycle" of swaps to occur, and moreover, that this is both a necessary and sufficient condition for optimality. It is natural to conjecture that a similar result occurs in the present setup.

(3) **Multiple marginals:** A natural generalization of the Skorokhod embedding problem is to consider the case where a sequence of measures, $\mu_1, \mu_2, \ldots, \mu_n$ are given, and the aim is to find a sequence of stopping times $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_n$ such that $B_{\tau_k} \sim \mu_k$, and such that the chosen sequence of stopping times minimizes $\mathbb{E}[\gamma((B_t)_{t \leq \tau_n}, \tau_1, \ldots, \tau_n)]$ for a suitable functional γ. In this setup, it is natural to ask whether there exists a suitable monotonicity principle, corresponding to Theorem 1.2.

(4) **Constrained embedding problems:** In this paper, we consider classical embedding problems, where the optimization is carried out over the class of solutions to (SEP). However, in many natural applications, one needs to further consider the class of *constrained* embedding problems: for example, where one minimizes some functional over the class of embeddings which also satisfy a restriction on the probability of stopping after a given time. It would be natural to derive generalizations of our duality results, and a corresponding monotonicity principle for such problems.

1.5. **Background.** Since the first solution to (SEP) by Skorokhod [54] the embedding problem has received frequent attention in the literature, with new solutions appearing regularly, and exploiting a number of different mathematical tools. Many of these solutions also prove to be, by design or accident, solutions of (OptSEP) for a particular choice of γ, e.g. [46, 48, 4, 28, 56, 41]. The survey [39] is a comprehensive account of all the solutions to (SEP) up to 2004, and references many articles which use or develop solutions to the Skorokhod embedding problem. More recently, novel twists on the classical Skorokhod embedding problem have been investigated by: Last et. al. [33], who consider the closely

related problem of finding unbiased shifts of Brownian motion (and where there are also natural connections to optimal transport); Hirsch et. al. [24], who have used solutions to the Skorokhod embedding problem to construct Peacocks; and Gassiat et. al. [22], who have exploited particular properties of Root's solution to construct efficient numerical schemes for SDEs.

The Skorokhod embedding problem has also recently received substantial attention from the mathematical finance community. This goes back to an idea of Hobson [25]: through the Dambis-Dubins-Schwarz Theorem, the optimization problems (OptSEP) are related to the pricing of financial derivatives, and in particular to the problem of *model-risk*. We refer the reader to the survey article [26] for further details.

Recently there has been much interest in optimal transport problems where the transport plan must satisfy additional martingale constraints. Such problems arise naturally in the financial context, but are also of independent mathematical interest, for example — mirroring classical optimal transport — they have important consequences for the study of martingale inequalities (see e.g. [9, 23, 40]). The first papers to study such problems include [27, 7, 20, 16], and this field is commonly referred to as *martingale optimal transport*. The Skorokhod embedding problem has been considered in this context by Galichon et. al. in [20]; through a stochastic control problem they recover the Azéma-Yor solution of the Skorokhod embedding problem. Notably, their approach is very different from the one pursued in the present paper: the approach of this paper is instead to use an analogue of c-cyclical monotonicity from classical optimal transport in the martingale context.

1.6. **Organization of the Article.** In Section 2 we establish the Root and the Rost embedding as well as a family of new embeddings. In Section 3 we recall some required definitions and results from optimal transport. In Section 4 we consider randomized stopping times on the Wiener space and establish some basic properties. In Section 5 we develop a dual problem to (OptSEP) and prove our duality using classical duality results from optimal transport. In Section 6 we will finally establish Theorem 1.2 by combining the duality theory with Choquet's capacity theorem. In Section 7 we use our results to establish the known solutions to (OptSEP) as well as further embeddings. For readers who are mainly interested in these applications, we also summarize the necessary results from earlier sections here. In Section 8 we describe extensions to Feller processes under certain assumptions, which we are able to verify for a large class of processes.

2. Particular embeddings

In this section we explain how Theorem 1.2 can be used to derive particular solutions to the Skorokhod embedding problem, (SEP), using the optimization problem (OptSEP). For much of the paper, we will consider (SEP) for measures μ where $\int x^2 \mu(dx) < \infty$. This constraint can be weakened to require only the first moment to be finite, subject to the restriction that the stopping time is *minimal*: that is, if τ is a stopping time such that $B_\tau \sim \mu$, then for any other stopping time τ',

$$B_{\tau'} \sim \mu \text{ and } \tau' \leq \tau \text{ implies } \tau' = \tau \text{ a.s.} \tag{2.1}$$

In the case where μ has a second moment, minimality and $\mathbb{E}[\tau] < \infty$ are equivalent. The extension of our results to the more general case will be discussed in Section 8.

2.1. **The Root embedding.** We recall the definition of the Root embedding, τ_{Root}, from (1.1), and we wish to recover Root's result ([46]) from an optimization problem. Let $\gamma(f, t) = h(t)$, where $h : \mathbb{R}_+ \to \mathbb{R}$ is a strictly convex function such that

$$\inf\{\mathbb{E}[h(\tau)] : \tau \text{ solves (SEP)}\} \tag{2.2}$$

is well posed and pick a minimizer $\hat{\tau}$ of (2.2) by Theorem 1.1. Then we have:

Theorem 2.1. *There exists a barrier $\mathcal{R}$ such that $\hat{\tau} = \inf\{t \geq 0 : (t, B_t) \in \mathcal{R}\}$. In particular the Skorokhod embedding problem has a solution of barrier type* (1.1).

Proof. Pick, by Theorem 1.2, a *γ-monotone* set $\Gamma \subseteq S$ (see Definition 1.4) such that $\mathbb{P}(((B_t)_{t \leq \hat{\tau}}, \hat{\tau}) \in \Gamma) = 1$. By convexity of h, the set of stop-go pairs is given by (cf. the comment below Definition 1.3 and Figure 2)

$$\mathsf{SG} = \{((f, s), (g, t)) \in S : f(s) = g(t), t < s\}.$$

As Γ is γ-monotone, $(\Gamma^< \times \Gamma) \cap \mathsf{SG} = \emptyset$. Define a closed and an open barrier by

$$\mathcal{R}_{\text{CL}} := \{(s, x) : \exists (g, t) \in \Gamma, g(t) = x, t \leq s\},$$
$$\mathcal{R}_{\text{OP}} := \{(s, x) : \exists (g, t) \in \Gamma, g(t) = x, t < s\},$$

and denote the respective hitting times by τ_{CL} and τ_{OP}. We claim that $\tau_{\text{CL}} \leq \hat{\tau} \leq \tau_{\text{OP}}$ a.s.

Note that $\tau_{\text{CL}} \leq \hat{\tau}$ holds by definition of τ_{CL}. To show the other inequality pick ω satisfying $((B_t(\omega))_{t \leq \hat{\tau}(\omega)}, \hat{\tau}(\omega)) \in \Gamma$ and assume for contradiction that $\tau_{\text{OP}}(\omega) < \hat{\tau}(\omega)$. Then there exists $s < \hat{\tau}(\omega)$ such that $(s, B_s(\omega)) \in \mathcal{R}_{\text{OP}}$. By definition of the open barrier, this means that there is some $(g, t) \in \Gamma$ such that $t < s$ and $g(t) = B_s(\omega)$. But $(f, s) := ((B_u(\omega))_{u \leq s}, s) \in \Gamma^<$, hence $((f, s), (g, t)) \in \mathsf{SG} \cap (\Gamma^< \times \Gamma)$ which is the desired contradiction. We finally observe that $\tau_{\text{CL}} = \tau_{\text{OP}}$ by the Strong Markov property, and the fact that one-dimensional Brownian motion immediately returns to its starting point. □

A consequence of this proof is that (on a given stochastic basis) there exists exactly one solution[3] of the Skorokhod embedding problem which minimizes (2.2). Assume that minimizers τ_1 and τ_2 are given. Then we can use an independent coin-flip to define a new minimizer $\bar{\tau}$ which is with probability $1/2$ equal to τ_1 and with probability $1/2$ equal to τ_2. By Theorem 2.2, $\bar{\tau}$ is of barrier-type and hence $\tau_1 = \tau_2$.

Remark 2.2. The following argument, due to Loynes [34], can be used to argue that barriers are unique in the sense that if two barriers solve (SEP), then their hitting times must be equal. Suppose that $\mathcal{R}$ and $\mathcal{S}$ are both *closed* barriers which embed μ. Note that we can take the closed barriers without altering the stopping properties. Consider the barrier $\mathcal{R} \cup \mathcal{S}$: let $A \subseteq \Omega_{\mathcal{R}} := \{x : (t, x) \in \mathcal{S} \implies (t, x) \in \mathcal{R}\}$. Then $\mathbb{P}(B_{\tau_{\mathcal{R} \cup \mathcal{S}}} \in A) \leq \mathbb{P}(B_{\tau_{\mathcal{R}}} \in A) = \mu(A)$. Similarly, for $A' \subseteq \Omega_{\mathcal{S}} := \{x : (t, x) \in \mathcal{R} \implies (t, x) \in \mathcal{S}\}$, $\mathbb{P}(B_{\tau_{\mathcal{R} \cup \mathcal{S}}} \in A') \leq \mathbb{P}(B_{\tau_{\mathcal{S}}} \in A') = \mu(A')$. Since $\mu(\Omega_{\mathcal{R}} \cup \Omega_{\mathcal{S}}) = 1$, $\tau_{\mathcal{R} \cup \mathcal{S}}$ embeds μ.

It is known (see Monroe [36]) that, when μ has a second moment, the second condition in (SEP), $\mathbb{E}[\tau] < \infty$ is equivalent to minimality of the stopping time (recall (2.1)). It immediately follows from the argument above that if the barriers $\mathcal{R}$ and $\mathcal{S}$ solve (SEP), then $\tau_{\mathcal{R}} = \tau_{\mathcal{S}}$ a.s.

With minor modifications the argument of Loynes also applies to the Rost solution discussed below as well as to a number of further classical embeddings presented in Section 7 below.

In Section 7.3 we will prove generalizations of Theorem 2.1 which admit similar conclusions in $\mathbb{R}^d$ and for general initial distributions.

We also note that the above proof of Theorem 2.1 is based on a heuristic derivation of the optimality properties of the Root embedding given by Hobson in [26]. Indeed Hobson's approach was the starting point of the present paper.

2.2. **The Rost embedding.** A set $\mathcal{R} \subseteq \mathbb{R}_+ \times \mathbb{R}$ is an *inverse barrier* if $(s, x) \in \mathcal{R}$ and $s > t$ implies that $(t, x) \in \mathcal{R}$. It has been shown by Rost [48] that under the condition $\mu(\{0\}) = 0$ there exists an inverse barrier such that the corresponding hitting time (in the sense of (1.1)) solves the Skorokhod problem. It is not hard to see that without this condition some additional randomization is required. We derive this using an argument almost identical to the one above.

Let $\gamma(f, t) = h(t)$, where $h : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly *concave* function such that the minimization of $\mathbb{E}[h(\tau)]$ over all solutions to (SEP) is well posed. Pick, by Theorem 1.1, a minimizer $\hat{\tau}$. Then we have:

---

[3]This was first established in [48], together with the optimality property of Root's solution.

(A) The Rost construction
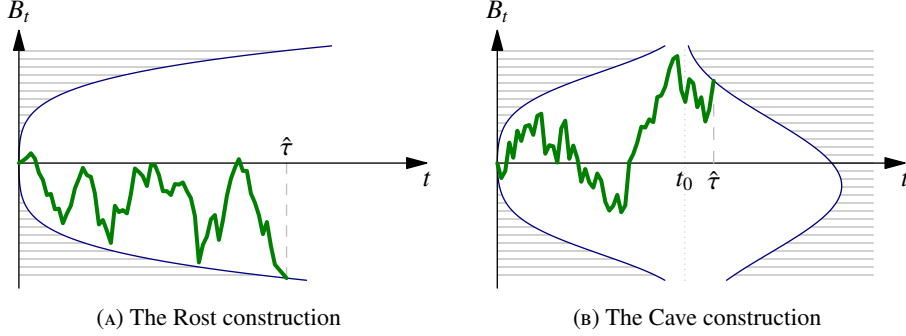

(B) The Cave construction

FIGURE 3. The barriers corresponding to the Rost and Cave embeddings

Theorem 2.3. *Suppose $\mu(\{0\}) = 0$. Then there exists an inverse barrier $\mathcal{R}$ such that $\hat{\tau} = \inf\{t \geq 0 : (t, B_t) \in \mathcal{R}\}$. In particular the Skorokhod embedding problem can be solved by the hitting time of an inverse barrier.*

Proof. Pick, by Theorem 1.2, a *γ-monotone* set $\Gamma \subseteq S$ such that $\mathbb{P}(((B_t)_{t \leq \hat{\tau}}, \hat{\tau}) \in \Gamma) = 1$. Note that due to the concavity of h the set of stop-go pairs is given by

$$\mathsf{SG} = \{((f, s), (g, t)) \in S : f(s) = g(t), s < t\}.$$

As Γ is γ-monotone, $(\Gamma^< \times \Gamma) \cap \mathsf{SG} = \emptyset$. Define open and closed inverse barriers by

$$\mathcal{R}_{\mathrm{OP}} := \{(s, x) : \exists (g, t) \in \Gamma, g(t) = x, s < t\},$$
$$\mathcal{R}_{\mathrm{CL}} := \{(s, x) : \exists (g, t) \in \Gamma, g(t) = x, s \leq t\},$$

and denote the respective hitting times by τ_{OP} and τ_{CL}. We claim that $\tau_{\mathrm{CL}} \leq \hat{\tau} \leq \tau_{\mathrm{OP}}$ a.s.

Note that $\tau_{\mathrm{CL}} \leq \hat{\tau}$ holds by definition of τ_{CL}. To show the other inequality pick ω satisfying $((B_t(\omega))_{t \leq \hat{\tau}(\omega)}, \hat{\tau}(\omega)) \in \Gamma$ and assume for contradiction that $\tau_{\mathrm{OP}}(\omega) < \hat{\tau}(\omega)$. Then there exists $s < \hat{\tau}(\omega)$ such that $(s, B_s(\omega)) \in \mathcal{R}_{\mathrm{OP}}$. By definition of the open inverse barrier, this means that there is some $(g, t) \in \Gamma$ such that $s < t$ and $g(t) = B_s(\omega)$. But $(f, s) := ((B_u(\omega))_{u \leq s}, s) \in \Gamma^<$, hence $((f, s), (g, t)) \in \mathsf{SG} \cap (\Gamma^< \times \Gamma)$ which is the desired contradiction.

Next we show that $\hat{\tau} = \tau_{\mathrm{OP}}$ a.s. Observe that we may assume that Γ contains no paths (f, s) such that $f(s) = 0$, since $\mu(\{0\}) = 0$ (if the original Γ contains points of this form, then we may replace Γ with $\tilde{\Gamma}$ which omits such paths without altering either the full support property, or the γ-monotone property). Then there exists an increasing function $b(t) = \inf\{x > 0 : (t, x) \in \mathcal{R}_{\mathrm{CL}}\}$ and a decreasing function $c(t) = \sup\{x < 0 : (t, x) \in \mathcal{R}_{\mathrm{CL}}\}$ such that $\tau_{\mathrm{CL}} = \inf\{t > 0 : B_t \notin (c(t), b(t))\}$, and $\tau_{\mathrm{OP}} = \inf\{t > 0 : B_t \notin [c(t), b(t)]\}$. In the case where $\lim_{t \to 0} b(t) > 0 > \lim_{t \to 0} c(t)$, then it is straightforward to show $\tau_{\mathrm{CL}} = \tau_{\mathrm{OP}}$ a.s., and the result follows.

Suppose that this is not the case. Since $\hat{\tau}$ embeds μ, and $\mu(\{0\}) = 0$, then $\hat{\tau} > 0$ a.s. In particular, $\hat{\tau} \leq \varepsilon$ can be made to have arbitrarily small probability, and in particular, for $\varepsilon' > 0$ sufficiently small, the laws of B_ε on $\{\hat{\tau} > \varepsilon, B_\varepsilon \in [c(\varepsilon) + \varepsilon', b(\varepsilon) - \varepsilon']\}$ and $\{\tau_{\mathrm{OP}} > \varepsilon, B_\varepsilon \in [c(\varepsilon) + \varepsilon', b(\varepsilon) - \varepsilon']\}$ can be made arbitrarily close in total variation norm, and with mass arbitrarily close to 1. By the same argument as above, on the intersection of these sets $\inf\{t > 0 : B_t \notin [c(t), b(t)]\} = \inf\{t > 0 : B_t \notin (c(t), b(t))\}$, and from a modified version of the argument that $\tau_{\mathrm{CL}} \leq \hat{\tau} \leq \tau_{\mathrm{OP}}$ we conclude that $\hat{\tau} = \tau_{\mathrm{OP}}$ when both $\hat{\tau}$ and τ_{OP} are greater than ε. Since ε is arbitrary, and $\hat{\tau} > 0$, then $\hat{\tau} = \tau_{\mathrm{OP}}$ a.s.   □

In Section 7.3 we will give a generalization of this result, which includes a more detailed argument for the final part of this proof. Note that it is also possible to show $\tau_{\mathrm{OP}} = \tau_{\mathrm{CL}}$: see the proof of Equation (2.9) in [13].

As in the case of the Root embedding we obtain that the maximizer of $\mathbb{E}[h(\tau)]$ is unique.

2.3. **The cave embedding.** In this section we give an example of a new embedding that can be derived from Theorem 1.2. It can be seen as a unification of the Root and Rost embeddings. A set $\mathcal{R} \subseteq \mathbb{R}_+ \times \mathbb{R}$ is a *cave barrier* if there exists $t_0 \in \mathbb{R}_+$, an inverse barrier $\mathcal{R}^0 \subseteq [0, t_0] \times \mathbb{R}$ and a barrier $\mathcal{R}^1 \subseteq [t_0, \infty) \times \mathbb{R}$ such that $\mathcal{R} = \mathcal{R}^0 \cup \mathcal{R}^1$. We will show that there exists a cave barrier such that the corresponding hitting time (in the sense of (1.1)) solves the Skorokhod problem. We derive this using an argument similar to the one above:

Fix $t_0 \in \mathbb{R}$ and pick a continuous function $\varphi : \mathbb{R}_+ \to [0, 1]$ such that

- $\varphi(0) = 0, \lim_{t\to\infty} \varphi(t) = 0, \varphi(t_0) = 1$
- φ is strictly concave on $[0, t_0]$
- φ is strictly convex on $[t_0, \infty)$.

It follows that φ is strictly increasing on $[0, t_0]$ and strictly decreasing on $[t_0, \infty)$. Let $\gamma((f, s)) = \varphi(s)$. Since φ is bounded, the problem to minimize $\mathbb{E}[\varphi(\tau)]$ over all solutions to (SEP) is well posed. Pick, by Theorem 1.1, a minimizer $\hat{\tau}$. Then we have:

Theorem 2.4 (Cave embedding). *Suppose $\mu(\{0\}) = 0$. Then there exists a cave barrier $\mathcal{R}$ such that $\hat{\tau} = \inf\{t \geq 0 : (t, B_t) \in \mathcal{R}\}$. In particular the Skorokhod embedding problem can be solved by a hitting time of a cave barrier.*

Proof. Pick, by Theorem 1.2, a γ-*monotone* set $\Gamma \subseteq S$ such that $\mathbb{P}(((B_t)_{t \leq \hat{\tau}}, \hat{\tau}) \in \Gamma) = 1$. The set of stop-go pairs is given by

$$\mathsf{SG} = \{((f, s), (g, t)) \in S \times S : f(s) = g(t); s < t \leq t_0 \text{ or } t_0 \leq t < s\}.$$

Indeed, for $s < t \leq t_0$ and any $(h, r) \in S$ we have

$$\gamma((f \oplus h, s + r)) + \gamma((g, t)) > \gamma((f, s)) + \gamma((g \oplus h, t + r))$$
$$\Leftrightarrow \quad \varphi(s + r) - \varphi(s) > \varphi(t + r) - \varphi(t)$$

which holds iff $t \mapsto \varphi(t + r) - \varphi(t)$ is strictly decreasing on $[0, t_0]$ for all $r > 0$. If $t + r, t \in [0, t_0]$ this follows from concavity of φ. In the case that $t \leq t_0, t + r > t_0$ this follows since φ' is strictly positive on $[0, t_0)$ and strictly negative on (t_0, ∞). The case $t_0 \leq t < s$ can be established similarly.

Then, we define an open cave barrier by

$$\mathcal{R}^0_{\mathrm{OP}} := \{(t, x) : \exists (f, s) \in \Gamma, t < s \leq t_0\}, \quad \mathcal{R}^1_{\mathrm{OP}} := \{(t, x) : \exists (f, s) \in \Gamma, t_0 \leq s < t\}$$

and $\mathcal{R}_{\mathrm{OP}} = \mathcal{R}^0_{\mathrm{OP}} \cup \mathcal{R}^1_{\mathrm{OP}}$ (resp. a closed cave barrier where we allow $t \leq s$ and $s \leq t$ in $\mathcal{R}^0_{\mathrm{CL}}$ and $\mathcal{R}^1_{\mathrm{CL}}$ resp.). We denote the corresponding hitting time by $\tau_{\mathcal{R}_{\mathrm{OP}}} = \tau_{R^0_{\mathrm{OP}}} \wedge \tau_{R^1_{\mathrm{OP}}}$ (resp. $\tau_{\mathcal{R}_{\mathrm{CL}}}$).

By the same argument as for the Root and Rost embedding it then follows that $\tau_{\mathcal{R}_{\mathrm{CL}}} \leq \hat{\tau} \leq \tau_{\mathcal{R}_{\mathrm{OP}}}$ a.s. and also that $\hat{\tau} = \tau_{\mathcal{R}_{\mathrm{OP}}}$ a.s., proving the claim. $\square$

2.4. **Remarks.** The arguments given here only use the properties of one-dimensional Brownian motion to show that our candidate stopping times τ_{OP} and $\hat{\tau}$ are the same. In Section 7.3 we will show that these arguments can be adapted to prove the existence of Rost and Root embeddings in a more general setting. In fact, in Sections 7 and 8 we will show that the above approach generalizes to a multi-dimensional setup and (sufficiently regular) Markov processes. In the case of the Root embedding it does not matter for the argument whether the starting distribution is a Dirac in 0 as in our setup or rather a more general distribution λ. For the Rost embedding a general starting distribution is slightly more difficult. In the case where λ and μ have common mass, then it may be the case that $\mathrm{proj}_{\mathbb{R}}(\mathcal{R}_{\mathrm{CL}} \cap (A \times \mathbb{R}_+)) = \{0\}$ for some set A — that is, all paths which stop at $x \in A$ do so at time zero. In this case it is possible that $\hat{\tau} < \tau_{\mathrm{OP}}$ when the process starts in A, and in general, some proportion of the paths starting on A must be stopped instantly. As a result, in the case of general starting measures, independent randomization is necessary. In the Rost case, it is also straightforward to compute the independent randomization which preserves the embedding property.

Other recent approaches to the Root and Rost embeddings can be found in [14, 38, 13, 14]. These papers largely exploit PDE techniques, and as a result, are able to produce

more explicit descriptions of the barriers, but the methods tend to be highly specific to the problem under consideration.

3. The classical Transport Problem

We will shortly review here some notions of transport theory which are used below or which will serve as motivation for analogous concepts in our probabilistic setup.

In abstract terms the transport problem (cf. [57, 58]) can be stated as follows: For probabilities λ, μ on Polish spaces X, Y the set $\mathsf{Cpl}(\lambda, \mu)$ of *transport plans* consists of all *couplings* between λ and μ. These are all measures on $X \times Y$ with X-marginal λ and Y-marginal μ. Associated to a *cost function* $c : X \times Y \to [0, \infty]$ and $\pi \in \mathsf{Cpl}(\lambda, \mu)$ are the *transport costs* $\int_{X \times Y} c(x, y) \, d\pi(x, y)$. The Monge-Kantorovich problem is to determine the value

$$\inf \left\{ \int c \, d\pi : \pi \in \mathsf{Cpl}(\lambda, \mu) \right\} \tag{3.1}$$

and to identify an *optimal* transport plan $\hat{\pi} \in \mathsf{Cpl}(\lambda, \mu)$, i.e. a minimizer of (3.1). Going back to Kantorovich, this is related to the following dual problem. Consider the set $\Phi(\lambda, \mu)$ of pairs (φ, ψ) of integrable functions $\varphi : X \to [-\infty, \infty)$ and $\psi : Y \to [-\infty, \infty)$ which satisfy $\varphi(x) + \psi(y) \leq c(x, y)$ for all $(x, y) \in X \times Y$. The dual counterpart of the Monge-Kantorovich problem is then to maximize

$$J(\varphi, \psi) = \int_X \varphi \, d\lambda + \int_Y \psi \, d\mu \tag{3.2}$$

over $(\varphi, \psi) \in \Phi(\lambda, \mu)$. In the literature duality has been established under various conditions, see for instance [58, Section 5] for a short overview.

Theorem 3.1 (Monge-Kantorovich Duality, [29, Theorem 2.2]). *Let $(X, \lambda), (Y, \mu)$ be Polish probability spaces and $c : X \times Y \to [0, \infty]$ be lower semi-continuous. Then*

$$\inf \left\{ \int c \, d\pi : \pi \in \mathsf{Cpl}(\lambda, \mu) \right\} = \sup \left\{ J(\varphi, \psi) : (\varphi, \psi) \in \Phi(\lambda, \mu) \right\}. \tag{3.3}$$

Moreover the duality relation pertains if the optimization in the dual problem is restricted to continuous and bounded functions φ, ψ.

A basic and important goal is to characterize minimizers through a tractable property of their support sets: a Borel set $\Gamma \subseteq X \times Y$ is *c-cyclically monotone* iff

$$c(x_1, y_2) - c(x_1, y_1) + \ldots + c(x_{n-1}, y_n) - c(x_{n-1}, y_{n-1}) + c(x_n, y_1) - c(x_n, y_n) \geq 0 \tag{3.4}$$

whenever $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in \Gamma$. A transport plan π is *c-cyclically monotone* if it assigns full measure to some cyclically monotone set Γ.

Concerning the origins of c-cyclical monotonicity in convex analysis and the study of the relation to optimality we mention [44, 31, 51, 21]. Intuitively speaking, c-cyclically monotone transport plans resist improvement by means of cyclical rerouting and optimal transport plans are expected to have this property. Indeed we have:

Theorem 3.2. *Let $c : X \times Y \to \mathbb{R}_+$ be a lower semi-continuous cost function. Then a transport plan is optimal if and only if it is c-cyclically monotone.*

Even in the case where c is the squared Euclidean distance this is a non trivial result, posed as an open question by Villani in [57, Problem 2.25]. Following contributions of Ambrosio and Pratelli [3], this problem was resolved by Pratelli [42] and Schachermayer and Teichmann [52] who established the clear-cut characterization stated in Theorem 3.2. Lower semi-continuity of the cost function can also be relaxed, as shown in [6] and [8].

We will need the following straightforward corollary of Theorem 3.1.

Corollary 3.3. *Let $\tilde{c} : X \times Y \times [0, t_0] \to \mathbb{R}$ be lower semi-continuous and bounded from below. Then*

$$\inf \left\{ \int \tilde{c} \, d\pi : \pi \in \mathcal{P}(X \times Y \times [0, t_0]), \mathsf{proj}_X(\pi) = \lambda, \mathsf{proj}_Y(\pi) = \mu \right\} \tag{3.5}$$

$$= \sup \left\{ J(\varphi, \psi) : (\varphi, \psi) \in L^\infty(\lambda) \times L^\infty(\mu), \varphi(x) + \psi(y) \leq \tilde{c}(x, y, t) \right\}. \tag{3.6}$$

Again, the duality relation pertains if the optimization in the dual problem is restricted to continuous and bounded functions φ, ψ.

4. Preliminaries on stopping times and filtrations

4.1. **Spaces and Filtrations.** In this section we mainly discuss the formal aspects of filtrations, measure theory, etc., and how classical notions relate to properties of functions on the space S introduced above.

We consider the space $\Omega = C(\mathbb{R}_+)$ of continuous paths with the topology of uniform convergence on compact sets. The elements of Ω will be denoted by ω. We denote the canonical process on Ω by $(B_t)_{t \geq 0}$, i.e. $B_t(\omega) = \omega_t$. As explained above we consider the set S of all continuous functions defined on some initial segment $[0, s]$ of $\mathbb{R}_+$; we will denote the elements of S by (f, s) and (g, t). The set S admits a natural partial ordering; we say that (g, t) extends (f, s) if $t \geq s$ and the restriction $g_{\upharpoonright [0,s]}$ of g to the interval $[0, s]$ equals f. In this case we write $(f, s) \prec (g, t)$. We consider S with the topology determined by the following metric: let $(f, s), (g, t) \in S$ and suppose $s \leq t$. We then say that (f, s) and (g, t) are ε-close if

$$d_S((f, s), (g, t)) := \max\left(t - s, \sup_{0 \leq u \leq s} |f(u) - g(u)|, \sup_{s \leq u \leq t} |g(u) - g(s)|\right) < \varepsilon. \quad (4.1)$$

Equipped with this topology, S is a Polish space.

For our arguments it will be important to be precise about the relationship between the sets $C(\mathbb{R}_+) \times \mathbb{R}_+$ and S. We therefore discuss the underlying filtrations in some detail.

We consider three different filtrations on the Wiener space $C(\mathbb{R}_+)$, the canonical or natural filtration $\mathcal{F}^0 = (\mathcal{F}_t^0)_{t \in \mathbb{R}_+}$, the right-continuous filtration $\mathcal{F}^+ = (\mathcal{F}_t^+)_{t \in \mathbb{R}_+}$, and the augmented filtration $\mathcal{F}^a = (\mathcal{F}_t^a)_{t \in \mathbb{R}_+}$ obtained from $(\mathcal{F}_t^0)_{t \in \mathbb{R}_+}$ by including all $\mathbb{W}$-null sets in $\mathcal{F}_0^0$. As Brownian motion is a continuous Feller process, $\mathcal{F}^a$ is automatically right-continuous, all $\mathcal{F}^a$-stopping times are predictable and all right-continuous $\mathcal{F}^a$-martingales are continuous. In particular, the $\mathcal{F}^a$-optional and the $\mathcal{F}^a$-predictable σ-algebras coincide (see e.g. [43, Corollary IV 5.7]). By [15, Thm. IV. 97, Rem. IV. 98] we also have that the $\mathcal{F}^0$-predictable,-optional and - progressive σ-algebras coincide because $\Omega = C(\mathbb{R}_+)$ is the set of *continuous* paths. Moreover, we will use the following result.

Theorem 4.1 ([15, Theorem IV. 78])**.** *Let $\mathcal{G}^a$ be the usual augmentation of the filtration $\mathcal{G}$. If τ is a predictable time w.r.t. $\mathcal{G}^a$, then there exists a predictable time τ' w.r.t. $\mathcal{G}$ such that $\tau = \tau'$ a.s. It follows that for every $\mathcal{G}^a$-predictable process $(X_t)_{t \in \mathbb{R}_+}$ there is an $\mathcal{G}$-predictable process $(X_t')_{t \in \mathbb{R}_+}$ which is indistinguishable from $(X_t)_{t \in \mathbb{R}_+}$.*

Of course, every $\mathcal{F}^a$-martingale has a continuous version. Not so commonly used but entirely straightforward is the following: if M is an $\mathcal{F}^0$-martingale then there is a version M' of M which is an $\mathcal{F}^0$-martingale and almost all paths of M' are continuous.

The message of Proposition 4.4 below is that a process $(X_t)_{t \in \mathbb{R}_+}$ is $\mathcal{F}^0$-predictable iff $(X_t)_{t \in \mathbb{R}_+}$ is $\mathcal{F}^0$-optional iff $X_t(\omega)$ can be calculated from the restriction $\omega_{\upharpoonright [0,t]}$. We introduce the mapping

$$r : C(\mathbb{R}_+) \times \mathbb{R}_+ \to S, \quad r(\omega, t) = (\omega_{\upharpoonright [0,t]}, t). \quad (4.2)$$

Note that the topology on S introduced in (4.1) coincides with the final topology induced by the mapping r; in particular r is a continuous open mapping. The mapping r is not a closed mapping: it is easy to see that there exist closed sets in $C(\mathbb{R}_+) \times \mathbb{R}_+$ with a non-closed image under r. However this does not happen for closed optional sets, see Proposition 4.4.

Remark 4.2. In the following we will say that $H : S \to \mathbb{R}$ is *continuous / right-continuous / etc.* if the corresponding property holds for the process $H \circ r$. Similarly we say that $H_1, H_2 : S \to \mathbb{R}$ are *indistinguishable* if this holds for the processes $H_1 \circ r, H_2 \circ r$ w.r.t. Wiener measure.

Definition 4.3. *We say that a process X is S- continuous if there exists a continuous function $h : S \to \mathbb{R}$ such that*

$$X_t(\omega) = h((\omega_{\restriction[0,t]}, t))$$

for all $t \geq 0$, $\mathbb{W}$-a.s.

It is trivially true that an S- continuous process is $\mathcal{F}^0$-adapted, and continuous ($\mathbb{W}$-a.s.). The converse is not generally true — consider the case where X_t is the local time of the Brownian motion at a level x. This is a continuous, $\mathcal{F}^0$-adapted process, however the corresponding function h is not a continuous mapping from S to $\mathbb{R}$. (Indeed, any path which has strictly positive local time can be approximated uniformly by paths with both zero and infinite local time).

Proposition 4.4. *$\mathcal{F}^0$-optional sets and functions on $C(\mathbb{R}_+) \times \mathbb{R}_+$ correspond to Borel-measurable sets and functions on S. More precisely we have:*

(1) A set $D \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+$ is $\mathcal{F}^0$-optional iff there is a Borel set $A \subseteq S$ with $D = r^{-1}(A)$.
(2) A process $X = (X_t)_{t \in \mathbb{R}_+}$ is $\mathcal{F}^0$-optional iff there is a Borel measurable $H : S \to \mathbb{R}$ such that $X = H \circ r$.

An $\mathcal{F}^0$-optional set $A \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+$ is closed in $C(\mathbb{R}_+) \times \mathbb{R}_+$ iff the corresponding set $r(A)$ is closed in S. An $\mathcal{F}^0$-optional process $X = H \circ r$ is S- continuous iff $H : S \to \mathbb{R}$ is continuous.

Remark 4.5. For $H : S \to \mathbb{R}$ we will slightly abuse notation and write $H(\omega, t)$ for $H \circ r(\omega, t)$.

For the proof of Proposition 4.4 we need another result from [15]. Write $a_t : \Omega \to \Omega$ for the stopping operation, i.e. $a_t(\omega)$ is the path which agrees with ω until t and stays constant afterwards.

Theorem 4.6 (cf. [15, Theorem IV. 97])**.** *Let $Z = (Z_t)_{t \in \mathbb{R}_+}$ be a measurable process on $\Omega = C(\mathbb{R}_+)$. Then Z is $\mathcal{F}^0$-optional iff $Z_t = Z_t \circ a_t$ for all $t \in \mathbb{R}_+$.*

Proof of Proposition 4.4. We will only prove the second assertion; the first one being an obvious consequence.

Set $\Omega' = \Omega \times \mathbb{R}_+$ and $a'(\omega, t) = (a_t(\omega), t)$. Then $Z_t = Z_t \circ a_t$ for all $t \in \mathbb{R}_+$ is equivalent to asserting that $Z = Z \circ a'$. Let S' be the set of all $(\omega, t) \in \Omega'$ for which $a'(\omega, t) = (\omega, t)$ (i.e. ω remains constant from t on). Note that r is a homeomorphism from S' to S and denote its inverse by r^{-1}.

Assume now that Z is an optional process. Then $Z = Z \circ a'$. Since $r = r \circ a'$ we have $Z = Z \circ r^{-1} \circ r \circ a' = (Z \circ r^{-1}) \circ r$. Hence we may take $H = Z \circ r^{-1}$ in Proposition 4.4.

Conversely, if $Z = H \circ r$, then we have $Z \circ a' = H \circ r \circ a' = H \circ r = Z$. Hence Z is optional.

The last assertion of the proposition follows from the identification of S with S'. $\quad\square$

Definition 4.7. *We call a set $D \subseteq S$ right complete if $(g, t) \in D$ and $(g, t) \prec (f, s)$ implies $(f, s) \in D$. We say $D \subseteq S$ is left complete if $(g, t) \in D$ and $(g, t) \succ (f, s)$ implies $(f, s) \in D$.*

Subsequently we will be interested in the *stochastic intervals* $[\![0, \tau]\!]$ for stopping times τ. In particular, recall that $[\![0, \tau]\!] = \{(\omega, t) : t \in [0, \tau(\omega)]\} \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+$. The following lemma connects characterizations of stopping times, sets in S, and stochastic intervals.

Lemma 4.8.   *(1) Suppose τ is an $\mathcal{F}^0$-stopping time. Then the set $D = r([\![\tau, \infty[\![) \subseteq S$ satisfies*
      (a) D is Borel and right complete;
      (b) if $(f, s) \in D$, the set $\{t : (f_{\restriction[0,t]}, t) \in D\}$ has a smallest element.
    Moreover, given such a set D, there exists an $\mathcal{F}^0$-stopping time τ determined by $[\![\tau, \infty[\![= r^{-1}(D)$.
    (2) Suppose τ is an $\mathcal{F}^+$-stopping time. Then the set $D = r(]\!]\tau, \infty[\![) \subseteq S$ satisfies
      (a) D is Borel and right complete;
      (b) if $(f, s) \in D$, the set $\{t : (f_{\restriction[0,t]}, t) \in D\}$ has no smallest element.

Moreover, given such a set D, there exists an $\mathcal{F}^+$-stopping time τ determined by $[\![\tau, \infty[\![= r^{-1}(D)$.

Proof. (1) First observe that if we set $\tau(\omega) = \inf\{t \geq 0 : (\omega_{\restriction[0,t]}, t) \in D\}$, it follows that τ is the required $\mathcal{F}^0$-stopping time. On the other hand, if τ is a $\mathcal{F}^0$-stopping time, then D is Borel (by Proposition 4.4), since $[\![\tau, \infty[\![$ is an optional set, and the other properties are straightforward.

(2) Observe that if τ is a $\mathcal{F}^+$-stopping time, then $\tau_n = \tau + 1/n$ is a sequence of strictly decreasing $\mathcal{F}^0$-stopping times, and $[\![\tau, \infty[\![= \bigcup_{n \in \mathbb{N}} [\![\tau_n, \infty[\![$. The conclusions follow from (1). $\qquad\square$

By Proposition 4.4 we then have:

Corollary 4.9. *The map r leaves stochastic intervals of $\mathcal{F}^+$-stopping times invariant, i.e. for every $\mathcal{F}^+$-stopping time κ it holds that $r^{-1}(r([\![0, \kappa]\!])) = [\![0, \kappa]\!]$. If κ is an $\mathcal{F}^0$-stopping time then also $r^{-1}(r([\![0, \kappa[\![)) = [\![0, \kappa[\![$.*

Recalling Definition 4.3, we call a martingale $(X_t)_{t \in \mathbb{R}_+}$ an *S- continuous martingale* if it can be written as $X_t(\omega) = h((\omega_{\restriction[0,t]}, t))$ for some $h : S \to \mathbb{R}$, which is continuous.

Definition 4.10. *Let $X : C(\mathbb{R}_+) \to \mathbb{R}$ be a measurable function which is bounded or positive. Then we define $\mathbb{E}[X|\mathcal{F}_t^0]$ to be the unique $\mathcal{F}_t^0$-measurable function satisfying*

$$\mathbb{E}[X|\mathcal{F}_t^0](\omega) = \int X((\omega_{\restriction[0,t]}) \oplus \omega') \, d\mathbb{W}(\omega').$$

Proposition 4.11. *Let $X \in C_b(C(\mathbb{R}_+))$. Then $X_t(\omega) := \mathbb{E}[X|\mathcal{F}_t^0](\omega)$ defines an S- continuous martingale. We denote this martingale by X^M.*

Proof. Note that $(f_n, s_n) \to (f, s)$ in S implies $f_n \oplus \omega \to f \oplus \omega$ in $C(\mathbb{R}_+)$ for every $\omega \in C(\mathbb{R}_+), \omega(0) = 0$, where $f \oplus \omega$ denotes concatenation of paths as usual. Hence, putting $X_g(\omega) := X(g \oplus \omega)$ the convergence $(f_n, s_n) \to (f, s)$ implies the pointwise convergence $X_{f_n}(\omega) \to X_f(\omega)$ for all $\omega \in C(\mathbb{R}_+)$ by continuity of X. Moreover, for $(f, s) \in S$

$$\int X_f(\omega) \, \mathbb{W}(d\omega) =: X^M(f, s)$$

is a function of (f, s). Since X is bounded, this allows to deduce using the dominated convergence theorem that

$$X^M(f_n, s_n) \to X^M(f, s).$$

This means that X^M is continuous on S, hence, S- continuous. $\qquad\square$

Proposition 4.12. *Suppose X is a bounded lower semi-continuous function on S. Then there exists a continuous martingale ψ such that $X_\tau^M = \psi_\tau$ almost surely for every $\mathcal{F}^+$-stopping time τ.*

Proof. Since X is lower semi-continuous, we can approximate from below by (bounded) continuous functions. In particular, let $\varphi^n \uparrow X$, and then the corresponding martingales $\varphi^{M,n}$ are S- continuous. In addition, we know that there exists a version of the martingale $(\mathbb{E}[X|\mathcal{F}_t^0])_{t \in \mathbb{R}_+}$ denoted by $(\psi_t)_{t \in \mathbb{R}_+}$ whose paths are almost surely continuous. It follows that $\varphi_\tau^{M,n} \uparrow \psi_\tau$ almost surely, and the claimed result holds. $\qquad\square$

4.2. **Approximation by particular stopping times.**

Lemma 4.13. *Let τ be an $\mathcal{F}^+$-stopping time. For any $\varepsilon > 0$ there is an $\mathcal{F}^+$-stopping time ρ such that*

(1) $\rho \leq \tau$
(2) $\mathbb{W}(\tau - \rho \geq \varepsilon) \leq \varepsilon$
(3) $[\![0, \rho]\!]$ is closed in $C(\mathbb{R}_+) \times \mathbb{R}_+$ and S.

Recall from Lemma 4.8 that for an $\mathcal{F}^+$-stopping time, the stochastic interval $[\![0, \rho]\!]$ can be identified with the Borel subset $r([\![0, \rho]\!])$ of S and from Proposition 4.4 that $[\![0, \rho]\!]$ is closed in $C(\mathbb{R}_+) \times \mathbb{R}$ iff $r([\![0, \rho]\!])$ is closed in S. It is also straightforward to see that the interval $[\![0, \rho]\!]$ is closed iff the function $\rho : C(\mathbb{R}_+) \to \mathbb{R}_+$ is upper semi-continuous.

Proof. Fix $\varepsilon > 0$. Assume first that

$$\tau(\omega) = \begin{cases} t & \omega \in A \\ \infty & \text{otherwise} \end{cases},$$

for some $\mathcal{F}_t^0$-measurable set A. If $t = 0$ we are done, so we assume $t > 0$. By Proposition 4.4, there is a Borel set $A_0 \subseteq C([0,t])$ such that $A = A_0 \oplus C((t,\infty))$. In this proof we will use that $\mathcal{F}_t^0$-measurable events can be identified with measurable subsets of $C([0,t])$. In particular, we will loosely write $\mathbb{W}(D)$ instead of $\mathbb{W}_{\restriction C([0,t])}(D)$ for some measurable $D \subseteq C([0,t])$. By outer regularity of $\mathbb{W}$ there is an open set $O \subseteq C[0,t], O \supseteq A_0$ such that $\mathbb{W}(O \setminus A_0) \leq \varepsilon$. Set

$$\rho(\omega) = \begin{cases} t & \omega \in O \\ \infty & \text{otherwise} \end{cases}.$$

Hence, ρ is an $\mathcal{F}^+$-stopping time and $[\![0, \rho]\!]$ is closed. By construction we have $\rho \leq \tau$ and

$$\mathbb{W}(\{|\tau - \rho| > \varepsilon\}) = \mathbb{W}(\{\tau = \infty, \rho < \infty\}) = \mathbb{W}(O \setminus A_0) \leq \varepsilon.$$

This proves the Lemma if τ is an $\mathcal{F}^0$-stopping time which only takes the values t and ∞.

Given an arbitrary $\mathcal{F}^+$-stopping time τ, there exists a sequence of $\mathcal{F}^0$-stopping times τ_n, taking only the values t_n and ∞, such that $\tau = \inf_n \tau_n$. Pick stopping times $\rho_n \leq \tau_n$ such that $\mathbb{W}(\rho_n + \varepsilon \leq \tau_n) < \varepsilon 2^{-n}$. Then $\rho := \inf_n \rho_n$ is still upper semicontinuous and $\rho \leq \tau, \mathbb{W}(\rho + \varepsilon \leq \tau) < \varepsilon$ as required. $\qquad\square$

Remark 4.14. To prove the previous lemma for a general starting distribution λ we need to make an additional approximation step at the start of the proof, when $t = 0$, that is for stopping times of the form

$$\tau(\omega) = \begin{cases} 0 & \omega_0 \in A \\ \infty & \text{else} \end{cases}.$$

In this case, take an open set $O \supseteq A$ with $\lambda(O \setminus A) \leq \varepsilon$. The rest of the argument stays the same.

Corollary 4.15. *Let τ be an $\mathcal{F}^+$-stopping time. Then there is a sequence of $\mathcal{F}^+$-stopping times τ_n such that*

(1) $\tau_n \uparrow \tau$ $\mathbb{W}$-*a.s.*
(2) $\mathbb{W}(\{\tau = \infty\} \cap \{\tau_n < \infty\}) \to 0$.
(3) *For each n the stochastic interval $[\![0, \tau_n]\!]$ is closed in $C(\mathbb{R}_+) \times \mathbb{R}_+$ and S.*

Proof. For each n apply the previous lemma with $\varepsilon_n = 2^{-n}$. $\qquad\square$

If τ is an $\mathcal{F}^a$-stopping time then the result still applies with a minor modification: we have to allow for an exceptional null set N.

4.3. **Randomized stopping times.** Working on the path space $C(\mathbb{R}_+)$, a stopping time τ is a mapping which assigns to each path ω the time $\tau(\omega)$ at which the path is stopped. If the stopping time depends on external randomization, then we may consider a path ω which is not stopped at a single point $\tau(\omega)$, but rather that there is a sub-probability measure ξ_ω on $\mathbb{R}$ which represents the probability that the path ω is stopped at a given time, conditional on observing the path ω. The aim of this section is to make this idea precise, and to establish connections with related properties in the literature. Specifically, the notion of a *randomized stopping time* has been established previously in [5, 35], and is closely connected to the class of *pseudo-stopping times*, which we will also exploit.

We consider the space

$$\mathsf{M} := \{\xi \in \mathcal{P}^{\leq 1}(C(\mathbb{R}_+) \times \mathbb{R}_+) : \xi(d\omega, dt) = \xi_\omega(dt)\mathbb{W}(d\omega), \xi_\omega \in \mathcal{P}^{\leq 1}(\mathbb{R}_+) \text{ for } \mathbb{W}\text{-a.e. } \omega\},$$

where $(\xi_\omega)_{\omega \in \Omega}$ is a disintegration of ξ in the first coordinate $\omega \in \Omega$. We equip M with the weak topology induced by the continuous bounded functions on $C(\mathbb{R}_+) \times \mathbb{R}_+$.

Recall that our principle interest is in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = C(\mathbb{R}_+)$ and $\mathbb{P} = \mathbb{W}$. Sometimes we will also consider the associated, right-continuous and complete filtration $(\mathcal{F}_t^a)_{t \geq 0}$. In what follows, we will also use a natural extension of the filtered probability space denoted $(\bar{\Omega}, \bar{\mathcal{F}}, (\bar{\mathcal{F}}_t)_{t \geq 0}, \bar{\mathbb{P}})$, where we take $\bar{\Omega} = \Omega \times [0, 1]$, $\bar{\mathcal{F}} = \mathcal{F} \otimes \mathcal{B}([0, 1])$, $\bar{\mathbb{P}}(A_1 \times A_2) = \mathbb{P}(A_1)\mathcal{L}(A_2)$, and set $\bar{\mathcal{F}}_t = \mathcal{F}_t^a \otimes \sigma([0, 1])$. Here, $\mathcal{L}$ denotes Lebesgue measure. We will write $\bar{B} = (\bar{B}_t)_{0 \leq t}$ for the canonical process on $\bar{\Omega}$, i.e. $\bar{B}_t(\omega, u) = \omega_t$.

We have the following result characterizing the class of *randomized stopping times*.

Theorem 4.16. *Let $\xi \in \mathsf{M}$. Then the following are equivalent:*

(1) *There is a Borel function $H : S \to [0, 1]$ such that H is right-continuous, decreasing and*

$$\xi_\omega([0, s]) := 1 - H(\omega_{\upharpoonright[0,s]}) \tag{4.3}$$

defines a disintegration of ξ w.r.t. to $\mathbb{W}$.

(2) *For every disintegration $(\xi_\omega)_{\omega \in \Omega}$ of ξ, for all $t \in \mathbb{R}_+$ and every Borel set $A \subseteq [0, t]$ the random variable*

$$X_t(\omega) = \xi_\omega(A)$$

is $\mathcal{F}_t^a$-measurable.

(3) *There is a disintegration $(\xi_\omega)_{\omega \in \Omega}$ of ξ such that for all $t \in \mathbb{R}_+$ and all $f \in C_b(\mathbb{R}_+)$ such that the support of f lies in $[0, t]$ the random variable*

$$X_t(\omega) = \xi_\omega(f)$$

is $\mathcal{F}_t^0$-measurable.

(4) *On the probability space $(\bar{\Omega}, \bar{\mathcal{F}}, (\bar{\mathcal{F}}_t)_{t \geq 0}, \bar{\mathbb{P}})$, the random time*

$$\rho(\omega, u) = \inf\{t \geq 0 : \xi_\omega([0, t]) \geq u\} \tag{4.4}$$

defines an $\bar{\mathcal{F}}$-stopping time.

Definition 4.17. *A measure $\xi \in \mathsf{M}$ satisfying the conditions of Theorem 4.16 is called a* randomized stopping time. *We denote the set of randomized stopping times by* RST.

Proof of Theorem 4.16. We start by establishing the equivalence of (1) and (4). It is straightforward to deduce (4) from (1). To show the other direction we can apply Theorem 4.1 to get an $\bar{\mathcal{F}}^0$-stopping time ρ' with $\mathbb{P}[\rho = \rho'] = 1$ where $\bar{\mathcal{F}}_t^0 = \mathcal{F}_t^0 \otimes \sigma([0, 1])$. Then, we set

$$1 - H(\omega_{\upharpoonright[0,t]}) := \int_0^1 \mathbb{1}_{[0,t]}(\rho'(\omega, u)) \, du$$

which is càdlàg and $\mathcal{F}^0$-optional. (1) follows by an application of Proposition 4.4.

Next, we show the equivalence of (1) and (2). We first establish that (1) implies (2). Let ξ_ω and ξ_ω' be disintegrations of ξ. From (1) it follows that there is some function H such that $\xi_\omega([0, s]) = 1 - H(\omega_{\upharpoonright[0,s]})$, and $t \mapsto 1 - H(\omega_{\upharpoonright[0,t]})$ is an increasing, càdlàg function for fixed ω. It follows that $\xi_\omega(A) = \int_A d(1 - H(\omega_{\upharpoonright[0,s]}))$ is $\mathcal{F}_t^0$-measurable, and hence, since $\xi_\omega'(A) = \xi_\omega(A)$ $\mathbb{W}$-a.s., $\xi_\omega'(A)$ is $\mathcal{F}_t^a$-measurable. To show that (2) implies (1) it is straightforward to show that (2) implies (4). Then the conclusion follows from the first part of the proof. Similar arguments establish the equivalence of (1) and (3). □

Remark 4.18. (1) The function H in (4.3) is unique up to indistinguishability (cf. Remark 4.2). We will designate this function H^ξ in the following. This function has a natural interpretation. $H^\xi(f, s)$ is the probability that a particle is still alive at time s given that it has followed the path f. We call H^ξ the *survival function* associated to ξ.

(2) We will say ξ is a non-randomized stopping time iff there is a disintegration $(\xi_\omega)_{\omega \in \Omega}$ of ξ such that ξ_ω is a Dirac-measure (of mass 1) for every ω. Clearly this means that $\xi_\omega = \delta_{\tau(\omega)}$ a.s. for some (non-randomized) stopping time τ. ξ is a non-randomized stopping time iff there is a version of H^ξ which only attains the values 0 and 1.

Corollary 4.19. *The set* RST *is closed.*

Proof. We consider condition (2) resp. (3) in Theorem 4.16; the goal is to express measurability of $X_t(\omega) := \xi_\omega(f)$, supp $f \subseteq [0, t]$ in a different fashion. Note that a bounded Borel function h is $\mathcal{F}_t^0$-measurable iff for all bounded Borel functions g

$$\mathbb{E}[hg] = \mathbb{E}[h\mathbb{E}[g|\mathcal{F}_t^0]],$$

of course this does not rely on our particular setup. By a functional monotone class argument, for $\mathcal{F}_t^0$-measurability of X_t it is sufficient to check that

$$\mathbb{E}[X_t(g - \mathbb{E}[g|\mathcal{F}_t^0])] = 0 \qquad (4.5)$$

for all $g \in C_b(C(\mathbb{R}_+))$. In terms of ξ, (4.5) amounts to

$$0 = \mathbb{E}[X_t(g - \mathbb{E}[g|\mathcal{F}_t^0])] = \int \mathbb{W}(d\omega) \int \xi_\omega(ds) f(s)(g - \mathbb{E}[g|\mathcal{F}_t^0])(\omega)$$
$$= \int f(s)(g - \mathbb{E}[g|\mathcal{F}_t^0])(\omega)\, \xi(d\omega, ds)$$

which is a closed condition by Proposition 4.11. □

Definition 4.20. *A randomized stopping time is* finite *iff* $\xi(C(\mathbb{R}_+) \times \mathbb{R}_+) = 1$. *The set of all finite randomized stopping times will be denoted by* RST^1.

Recall from (1.5) that $\Gamma^< = \{(f, s) : \exists (g, t) \in \Gamma, s < t, (f, s) = (g_{\restriction [0,s]}, s)\}$ for $\Gamma \subseteq S$. We denote the push forward of a measure α by a function F by $F(\alpha)$.

Lemma 4.21. *Let* $\xi \in$ RST^1. *Then there exists a Borel set* $\Gamma \subseteq S$ *with* $r(\xi)(\Gamma) = 1$ *and* $\Gamma^< \cap \Gamma = \emptyset$ *iff* $\xi = \delta_\tau$ *for some* $\mathcal{F}^a$-*stopping time* τ.

Proof. Let τ be an $\mathcal{F}^a$-stopping time. By Theorem 4.1, there exists an $\mathcal{F}^0$-stopping time τ' with $\tau = \tau'$ $\mathbb{W}$-a.s. Then $\Gamma = \{r(\omega, \tau'(\omega)), \omega \in \Omega\}$ satisfies $\Gamma^< \cap \Gamma = \emptyset$ and $\xi = \delta_\tau$ is concentrated on Γ. Here Γ is an analytic set and hence universally measurable. We may thus replace Γ with a Borel subset of full ξ-measure to obtain the desired conclusion.

Pick $\xi \in$ RST^1 and a set Γ on which ξ is concentrated. $\Gamma^< \cap \Gamma = \emptyset$ implies that for any ω the set $\{t : r(\omega, t) \in \Gamma\}$ is at most single-valued. Put $D := \{(g, t) : \exists (f, s) \in \Gamma, (f, s) \prec (g, t)\}$. By Lemma 4.8 this defines an $\mathcal{F}^0$-stopping time on a subset of full measure (recall that ξ is only concentrated on Γ) proving the result. □

Given $\xi \in$ M and $s \in \mathbb{R}_+$ we define the measure $\xi \wedge s \in$ M to be the random time which is the minimum of ξ and s; formally this means that for $\omega \in \Omega$ and $A \subseteq \mathbb{R}_+$

$$(\xi \wedge s)_\omega(A) := \xi_\omega(A \cap [0, s)) + \delta_s(A)(1 - \xi_\omega([0, s))).$$

Assume that $(M_s)_{s \in \mathbb{R}_+}$ is a process on Ω. Then the stopped process $(M_s^\xi)_{s \in \mathbb{R}_+}$ is defined to be the *probability measure* on $\mathbb{R}$ such that for all bounded and measurable functions f

$$\int_{\mathbb{R}} f(x)\, M_s^\xi(dx) := \int f(M_t(\omega))\, (\xi \wedge s)(d\omega, dt) = \mathbb{E}_{\xi \wedge s}[f(M_t)].$$

Otherwise said M_s^ξ is the image measure of $\xi \wedge s$ under the map $M : C(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}, (\omega, t) \mapsto M_t(\omega)$. We write $\lim_{s \to \infty} M_s^\xi = M_\xi$ if it exists.

4.4. **Pseudo-randomized stopping times and dual optional projections.** We wish to characterize the subset of M corresponding to RST. A natural candidate for such a condition would be via the optional stopping theorem:

Definition 4.22. *Let* PRST *be the set of all* pseudo-randomized stopping times, *that is, the set of* $\xi \in$ M *satisfying*

$$\int X(\omega)\, \mathbb{W}(d\omega) = \int X_s^M(\omega)\, (\xi \wedge t)(d\omega, ds), \qquad (4.6)$$

for all $t \geq 0$ *and all* $X \in \mathcal{B}_b(C(\mathbb{R}_+))$, *the class of bounded Borel functions on* $C(\mathbb{R}_+)$.

Unfortunately RST is a proper subset of PRST; it is not hard to see this from [37]. By a functional monotone class argument it is sufficient to check (4.6) for all $X \in C_b(C(\mathbb{R}_+))$, in particular we have:

Proposition 4.23. *The set* PRST *is closed.*

Fortunately, the difference between RST and PRST is not seen by optional processes: given a pseudo-randomized stopping time ξ there always exists a randomized stopping time $\bar{\xi}$ such that for every optional bounded or positive process X we have $\int X \, d\xi = \int X \, d\bar{\xi}$.

Lemma 4.24. *Let* $\xi \in$ PRST. *Set* $A_t(\omega) := \xi_\omega([0, t])$. *Define* ξ^o *through* $\xi^o_\omega([0, t]) := A^o_t(\omega)$, *where* A^o *denotes the dual optional projection of* A. *Then* $\xi^o \in$ RST.

Proof. We prove this for a finite time ξ. By Theorem 4.16, we have to show that A^o_t is $\mathcal{F}^a_t$-measurable for every t, A^o is increasing nonnegative and bounded by 1. The only property that does not follow directly from the definition of dual optional projection is the boundedness by 1. As $\xi \in$ PRST we have using $X \equiv 1$

$$\mathbb{E}[A^o_\infty] = 1.$$

Hence, it is sufficient to show that $A^o_\infty \leq 1$. To this end, assume that $D := \{A^o_\infty(\omega) > 1\}$ has positive mass. Then we have using (4.6) and $X = 1_D$

$$\mathbb{W}(D) = \mathbb{E}[X] = \mathbb{E}[X^M_0] = \mathbb{E}_\xi[X^M] = \mathbb{E}\int_0^\infty X^M_s \, dA_s = \mathbb{E}\int_0^\infty X^M_s \, dA^o_s = \mathbb{E}XA^o_\infty > \mathbb{W}(D),$$

implying that $\mathbb{W}(D) = 0$. Hence, $\xi^o \in$ RST. $\qquad\square$

Clearly every pseudo-randomized stopping time $\xi \in$ PRST can be represented as a positive random variable on $\bar{\Omega}$ in a similar manner to (4.4) by taking $\rho(\omega, u) = \inf\{t \geq 0 : \xi_\omega([0, t]) \geq u\}$. The message of the above result is that, for any such ξ, and any optional bounded or positive process X on Ω, there exists a stopping time ρ^o on the extended space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathcal{F}}_t, \bar{\mathbb{P}})$ such that $\bar{\mathbb{E}}[\bar{X}_{\rho \wedge t}] = \mathbb{E}_{\xi \wedge t}[X_s] = \mathbb{E}_{\xi^o \wedge t}[X_s] = \bar{\mathbb{E}}[\bar{X}_{\rho^o \wedge t}]$ for any $t \geq 0$, where $\bar{X}(\omega, u) = X(\omega)$. Of course, we will eventually be interested in the subset of stopping times corresponding moreover to (SEP) — that is, we are specifically interested in the subset of PRST which both embed μ, and which satisfy a further natural criterion corresponding to the second condition in (SEP). However, by taking the optional processes $X_t = f(B_t)$ for bounded f, we immediately see that $\bar{B}_\rho \sim \bar{B}_{\rho^o}$, considering $X_t = t$ we obtain $\mathbb{E}_\xi[T] = \mathbb{E}_{\xi^o}[T] = \bar{\mathbb{E}}[\rho^o]$, where T denotes the projection

$$T : C(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}_+. \tag{4.7}$$

We observe also that to show that the process B^ξ_t is uniformly integrable, we need to show $\lim_{R \to \infty} \sup_t \int_{|x| > R} |x| \, B^\xi_t(dx) = 0$. However, with the above definitions, we have $B^\xi_t(dx) = B^{\xi^o}_t(dx)$, and so B^ξ is uniformly integrable if and only if B^{ξ^o} is also.

From now on we make the assumption that the measure μ which we want to embed has mean 0 and finite second moment[4]

$$V := \int x^2 \, \mu(dx) < \infty. \tag{4.8}$$

Then by the above arguments, and as a direct consequence of the same result for the stopping time ρ^o, we have:

Lemma 4.25. *Let* $\xi \in$ PRST. *Assume that* $B_\xi = \mu$, *i.e.* $\bar{B}_\rho \sim \mu$, *where* ρ *is the random time on* $\bar{B}$ *corresponding to* ξ. *Then the following are equivalent:*

(1) $\bar{\mathbb{E}}[\rho] < \infty$,
(2) $\bar{\mathbb{E}}[\rho] = V$,
(3) $(\bar{B}_{\rho \wedge t})$ *is uniformly integrable.*

Definition 4.26. *We denote by* PRST(μ) *the set of all pseudo-randomized stopping times satisfying the conditions in Lemma 4.25. Similarly, we define* RST$(\mu) =$ PRST$(\mu) \cap$ RST.

---

[4]This assumption is only made for ease of exposition. We refer to Section 8 and in particular Proposition 8.3 for the general case.

An immediate consequence is:

Corollary 4.27. *Let X_t be an optional process. Then for every $\xi \in \mathsf{PRST}(\mu)$, there exists $\xi^o \in \mathsf{RST}(\mu)$ with $\mathbb{E}_\xi[X_t] = \mathbb{E}_{\xi^o}[X_t]$.*

The main reason why we consider *randomized* stopping times and their pseudo-randomized counterparts is that they have the following property:

Theorem 4.28. *The set $\mathsf{PRST}(\mu)$ is compact.*

Proof. By Prohorov's theorem we have to show that $\mathsf{PRST}(\mu)$ is tight and that $\mathsf{PRST}(\mu)$ is closed.

Tightness. Fix $\varepsilon > 0$ and take R such that $V/R \leq \varepsilon/2$. Then, for any $\xi \in \mathsf{PRST}(\mu)$ we have $\xi(T > R) \leq \varepsilon/2$. As $C(\mathbb{R}_+)$ is Polish there is a compact set $\tilde{K} \subseteq C(\mathbb{R}_+)$ such that $\mathbb{W}(\complement \tilde{K}) \leq \varepsilon/2$. Set $K := \tilde{K} \times [0, R]$. Then K is compact and we have for any $\xi \in \mathsf{PRST}(\mu)$

$$\xi(\complement K) \leq \mathbb{W}(\complement \tilde{K}) + \xi(T > R) \leq \varepsilon.$$

Hence, $\mathsf{PRST}(\mu)$ is tight.

Closedness. Take a sequence $(\xi_n)_{n\in\mathbb{N}}$ in $\mathsf{PRST}(\mu)$ converging to some ξ. Putting $h : C(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}, (\omega, t) \mapsto \omega(t)$ we have to show that $h(\xi) = \mu$ and that $\mathbb{E}_\xi[T] < \infty$. Note that h is a continuous map. Take any $g \in C_b(\mathbb{R})$. Then $g \circ h \in C_b(C(\mathbb{R}_+) \times \mathbb{R}_+)$. Thus, we have that

$$\int g \, d\mu = \lim_n \int_{C(\mathbb{R}_+)\times\mathbb{R}_+} g \circ h \, d\xi_n = \int_{C(\mathbb{R}_+)\times\mathbb{R}_+} g \circ h \, d\xi = \int g \, dh(\xi).$$

Hence, we have $h(\xi) = \mu$. Moreover, the set $\{(\omega, t) : t \leq L\}$ is closed. Hence, by the Portmanteau theorem, for any $L \geq 0$

$$\limsup \xi_n(t \leq L) \leq \xi(t \leq L).$$

This readily implies that $\mathbb{E}_\xi[T] \leq \liminf \mathbb{E}_{\xi_n}[T] = V < \infty.$ □

Since RST is a closed set we also have:

Corollary 4.29. *The set $\mathsf{RST}(\mu)$ of randomized stopping times which embed μ is compact.*

Our use of randomization to achieve compactness of a set of stopping times has similarities to the work of Baxter and Chacon [5]. However their setup is different, and their intended applications are not connected to Skorokhod embedding.

4.5. **Joinings / Tagged Stopping Times.** We now add another dimension: assume that (Y, ν) is some Polish probability space. The set of all *tagged pseudo-randomized stopping times* or rather *joinings* $\mathsf{JOIN}(\mathbb{W}, \nu) = \mathsf{JOIN}(\nu)$ is given by

$$\left\{ \pi \in \mathcal{P}^{\leq 1}(C(\mathbb{R}_+) \times \mathbb{R}_+ \times Y), \mathsf{proj}_{C(\mathbb{R}_+)\times\mathbb{R}_+}(\pi_{\restriction C(\mathbb{R}_+)\times\mathbb{R}_+\times B}) \in \mathsf{PRST}, B \in \mathcal{B}(Y), \mathsf{proj}_Y(\pi) \leq \nu \right\}.$$

We shall also write $\mathsf{JOIN}^1(\mathbb{W}, \nu)/\mathsf{JOIN}^1(\nu)$ for the subset of $\pi \in \mathsf{JOIN}(\nu)$ having mass 1.

Remark 4.30. Write pred for the σ-algebra of $\mathcal{F}^0$-predictable sets in $C(\mathbb{R}_+) \times \mathbb{R}_+$.

We call a set $A \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+ \times Y$ predictable if it is an element of $\mathsf{pred} \otimes \mathcal{B}(Y)$. We will say that a function defined on $C(\mathbb{R}_+) \times \mathbb{R}_+ \times Y$ is predictable if it is measurable w.r.t. $\mathsf{pred} \otimes \mathcal{B}(Y)$.

5. The Optimization Problem and Duality

5.1. **The Primal Problem.** As defined in (OptSEP) in the introduction, our primal problem is to minimize the value corresponding to a functional $\gamma : S \to \mathbb{R}$, where the minimization is taken over stopping times of Brownian motion defined on some probability space. Using the concepts presented in the previous section, we obtain an equivalent problem if we take B to be the canonical process on Wiener space $C(\mathbb{R}_+) = \Omega$ and minimize over all randomized stopping times, i.e. we have

$$P_\gamma = \inf \left\{ \int \gamma(\omega, t) \, \xi(d\omega, dt) : \xi \in \mathsf{RST}(\mu) \right\}. \tag{5.1}$$

We make the important comment that the optimization problem is not altered if the set $\mathsf{RST}(\mu)$ is replaced by $\mathsf{PRST}(\mu)$ (cf. Corollary 4.27).

In the following we will mainly work with the technically convenient formulation given in (5.1). It immediately allows us to establish the existence of optimal stopping times:

Proposition 5.1. *Assume that* $\gamma : S \to \mathbb{R}$ *is* S*- lower semi-continuous and bounded from below in the sense*[5] *that for some constants* $a, b, c \in \mathbb{R}_+$

$$-\left(a + bs + c \max_{r \le s} B_r^2\right) \le \gamma((B_r)_{r \le s}, s) \tag{5.2}$$

holds on $C(\mathbb{R}_+)$. *Then the functional*

$$\xi \mapsto \int_{C(\mathbb{R}_+) \times \mathbb{R}_+} \gamma(\omega, t)\, \xi(d\omega, dt) \tag{5.3}$$

is lower semi-continuous and in particular (5.1) *admits a minimizer.*

Since every randomized stopping time can be represented as a usual stopping time on an enlarged probability space, Theorem 1.1 is a particular consequence of this result.

Proof of Proposition 5.1. It is straightforward to see that the functional (5.3) is lower semi-continuous if $\gamma : S \to \mathbb{R}$ is S- lower semi-continuous and bounded from below by a constant. (This is spelled out in detail for instance in [58, Chapter 4] in the context of classical optimal transport.)

For the general case we recall the pathwise version of Doob's inequality (see [1])

$$\max_{r \le s} B_r^2 \le \underbrace{\int_0^s 4 \max_{t \le r} |B_t|\, dB_r}_{=:M_s} + 4B_s^2. \tag{5.4}$$

We emphasize that we can understand the integral defining M in a *pathwise* fashion, this is possible since $r \mapsto \max_{t \le r} |B_t|$ is increasing; we refer to [1] for details. In fact it is straightforward to show that M is an S- continuous martingale satisfying $|M_t| < 2 \max_{r \le t} B_r^2$. It follows that $\tilde{\gamma}(f, s) := \gamma(f, s) + bs + c(M(f, s) + 4f(s)^2)$ is bounded from below and hence $\xi \mapsto \int \tilde{\gamma}\, d\xi$ is lower semi-continuous. As the value of $\int bs + c(M_s(\omega) + 4B_s^2(\omega))\, d\xi(\omega, s)$ is the same for any $\xi \in \mathsf{RST}(\mu)$ the functional (5.3) is lower semi-continuous as well. $\square$

5.2. The dual problem.

Theorem 5.2. *Let* $\gamma : S \to \mathbb{R}$ *be* S*- lower semi-continuous and bounded from below in the sense of* (5.2). *Put*

$$D_\gamma = \sup \left\{ \int \psi(y)\, d\mu(y) : \psi \in C(\mathbb{R}), \exists \varphi, \begin{array}{l} \varphi \text{ is an } S\text{- continuous martingale}, \varphi_0 = 0 \\ \varphi_t(\omega) + \psi(\omega(t)) \le \gamma(\omega, t),\ (\omega, t) \in \Omega \times \mathbb{R}_+ \end{array} \right\}$$

where φ, ψ *satisfy* $|\varphi_t| \le a + bt + cB_t^2$, $|\psi(y)| \le a + by^2$ *for some* $a, b, c > 0$. *Then we have the duality relation*

$$P_\gamma = D_\gamma. \tag{5.5}$$

Theorem 5.2 has close analogues in the mathematical finance literature. In particular, using Hobson's time change argument ([25, 26]), Theorem 5.2 is comparable to the work of Dolinsky and Soner [18, 17]. Similar duality results in a discrete time framework are established by Bouchard and Nutz [9] among others.

Using the same argument as above, we see that it suffices to establish Theorem 5.2 in the case where γ is bounded from below. As usual, one part of the duality relation is straightforward to verify:

Lemma 5.3. *With the above notations and assumptions we have* $D_\gamma \le P_\gamma$.

---

[5]Other conditions which guarantee uniform integrability of the negative part of γ w.r.t. solutions of (SEP) would suffice as well.

Proof. Take (φ, ψ) satisfying the dual constraint and $\xi \in \mathsf{RST}(\mu)$. Then we have

$$\int \psi(y)\,\mu(dy) = \int \psi(\omega(t))\,\xi(d\omega, dt) + \int \varphi_t^M(\omega)\,\xi(d\omega, dt) \leq \int \gamma(\omega, t)\,\xi(d\omega, dt). \qquad \square$$

5.3. Transport formulation.

The strategy for the proof of Theorem 5.2 is to translate the embedding problem for μ into a (modified) transport problem between the Wiener measure $\mathbb{W}$ and the target distribution μ. To this end we equip the space $C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}$ with the cost function

$$c(\omega, t, y) := \begin{cases} \gamma(\omega, t) & \text{if } \omega(t) = y \\ \infty & \text{otherwise} \end{cases}. \tag{5.6}$$

The candidates for the corresponding primal and dual problem will be elements of the sets

$$\mathsf{JOIN}^{1,V}(\mathbb{W}, \mu) := \{\pi \in \mathsf{JOIN}^1(\mathbb{W}, \mu) : \mathbb{E}_\pi[T] \leq V\},$$

$$\mathsf{DC}_c^V := \left\{ (\varphi, \psi) : \begin{array}{l} \varphi \text{ is an } S\text{- continuous bounded martingale, } \psi \in C_b(\mathbb{R}),\ \exists \alpha \geq 0, \\ \varphi_t(\omega) + \psi(y) - \alpha(t - V) \leq c(\omega, t, y),\ \text{for all } \omega \in C(\mathbb{R}_+), y \in \mathbb{R}, t \in \mathbb{R}_+ \end{array} \right\},$$

where T denotes the projection onto $\mathbb{R}_+$, $V = \int x^2\,\mu(dx)$ and we used $Y = \mathbb{R}$ in the definition of $\mathsf{JOIN}^1(\mathbb{W}, \mu)$ (see Section 4.5). We then consider the optimization problems

$$\mathcal{P}_c^V := \inf_{\pi \in \mathsf{JOIN}^{1,V}(\mathbb{W}, \mu)} \int c(\omega, t, y)\,\pi(d\omega, dt, dy), \tag{5.7}$$

$$\mathcal{D}_c^V := \sup_{(\varphi, \psi) \in \mathsf{DC}^V(c)} \varphi_0 + \mu(\psi). \tag{5.8}$$

For the remainder of this section we will use the caligraphic letters $\mathcal{P}$ and $\mathcal{D}$ for (primal and dual, respectively) optimization problems on the space $C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}$.

Proposition 5.4. *With the above definitions and assumptions $\mathcal{D}_c^V \leq D_\gamma$ and $P_\gamma \leq \mathcal{P}_c^V$.*

Proof. Consider $p(\omega, t, y) := (\omega, t)$. If $\pi \in \mathsf{JOIN}^{1,V}(\mathbb{W}, \mu)$ satisfies $\int c\,d\pi < \infty$ it is concentrated on $\{(\omega, t, y) : \omega(t) = y\}$. Then, we have $\xi := p(\pi) \in \mathsf{PRST}(\mu)$ and $\int c\,d\pi = \int \gamma\,d\xi$. Hence, $P_\gamma \leq \mathcal{P}_c^V$.

It remains to show that $\mathcal{D}_c^V \leq D_\gamma$. A bounded pair (φ, ψ) belongs to DC_c^V iff there is $\alpha \geq 0$ such that for all $\omega \in \Omega, y \in \mathbb{R}, t \in \mathbb{R}_+$

$$\varphi_t(\omega) + \psi(y) - \alpha(t - V) \leq c(\omega, t, y)$$

which holds iff for all $\omega \in \Omega, t \in \mathbb{R}_+$

$$\varphi_t(\omega) + \psi(\omega(t)) - \alpha(t - V) \leq \gamma(\omega, t).$$

We rewrite this as

$$[\varphi_t(\omega) + \alpha(\omega(t)^2 - t)] + [\psi(\omega(t)) - \alpha\omega(t)^2 + \alpha V] \leq \gamma(\omega, t). \tag{5.9}$$

The alternative representation in (5.9) is useful to us since $\omega(t)^2 - t$ is an S- continuous martingale starting in 0. Setting

$$\bar{\varphi}_t(\omega) = \varphi_t(\omega) + \alpha(\omega(t)^2 - t) \quad \text{and} \quad \bar{\psi}(y) = \psi(y) - \alpha y^2 + \alpha V,$$

we have $\bar{\varphi}_t(\omega) + \bar{\psi}(\omega(t)) \leq \gamma(\omega, t)$. This means that $(\bar{\varphi} - \bar{\varphi}_0, \bar{\psi} + \bar{\varphi}_0)$ satisfy the constraint in the dual problem in (5.5). Recalling that V was defined by $V = \int y^2\,\mu(dy)$ we have $\int \bar{\psi}(y)\,\mu(dy) = \int \psi(y)\,\mu(dy)$. Therefore, we can conclude that

$$\mathcal{D}_c^V \leq D_\gamma. \qquad \square$$

To establish Theorem 5.2 we need to prove that $\mathcal{P}_c^V = \mathcal{D}_c^V$. Before we derive this in Proposition 5.8 below, we require a further, auxiliary duality result.

5.4. **A Non-Adapted (NA) Duality Result.** We consider the candidate sets

$$\mathsf{TM}^V(\mathbb{W}, \mu) := \{\pi \in \mathcal{P}(C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}) : \mathsf{proj}_{C(\mathbb{R}_+)}(\pi) = \mathbb{W}, \mathsf{proj}_{\mathbb{R}}(\pi) = \mu, \mathbb{E}_\pi[T] \leq V\},$$

$$\mathsf{DC}_c^{NA,V} := \left\{ (\varphi, \psi) \in C_b(\Omega) \times C_b(\mathbb{R}) : \exists \alpha \geq 0, \begin{array}{l} \varphi(\omega) + \psi(y) - \alpha(t - V) \leq c(\omega, t, y) \\ \text{for all } \omega \in \Omega, y \in \mathbb{R}, t \geq 0 \end{array} \right\}.$$

Note that the set TM^V is compact as a consequence of Prohorov's theorem. Corresponding to the above candidate sets we consider optimization problems

$$\mathcal{P}_c^{NA,V} := \inf_{\pi \in \mathsf{TM}^V(\mathbb{W}, \mu)} \int c \, d\pi, \tag{5.10}$$

$$\mathcal{D}_c^{NA,V} := \sup_{(\varphi, \psi) \in \mathsf{DC}_c^{NA,V}} \mathbb{W}(\varphi) + \mu(\psi). \tag{5.11}$$

Proposition 5.5. *Let $c : C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ be lower semi-continuous and bounded from below. Then*

$$\mathcal{D}_c^{NA,V} = \mathcal{P}_c^{NA,V}. \tag{5.12}$$

As before it is obvious that $\mathcal{D}_c^{NA,V} \leq \mathcal{P}_c^{NA,V}$ (cf. Lemma 5.3). To simplify the proof of the reverse inequality we note the following

Lemma 5.6. *If (5.12) is valid for a sequence of continuous bounded functions $c_n, n \geq 1$ such that $c_n \uparrow c$ then (5.12) applies also to c.*

Proof. We have to prove that $\mathcal{D}_c^{NA,V} \geq \mathcal{P}_c^{NA,V}$. For each k let $\pi_k \in \mathsf{TM}^V(\mathbb{W}, \mu)$ be such that

$$\mathcal{P}_{c_k}^{NA,V} \geq \int c_k \, d\pi_k - 1/k.$$

By compactness of $\mathsf{TM}^V(\mathbb{W}, \mu)$ there is a subsequence, still denoted by k, such that $(\pi_k)_{k \in \mathbb{N}}$ converges weakly to some $\pi \in \mathsf{TM}^V(\mathbb{W}, \mu)$. Then by monotone convergence using the monotonicity of the sequence $(c_k)_{k \in \mathbb{N}}$ we have

$$\mathcal{P}_c^{NA,V} \leq \int c \, d\pi = \lim_{m \to \infty} \int c_m \, d\pi = \lim_{m \to \infty} \left(\lim_{k \to \infty} \int c_m \, d\pi_k \right)$$

$$\leq \lim_{m \to \infty} \left(\lim_{k \to \infty} \int c_k \, d\pi_k \right) = \lim_{k \to \infty} \mathcal{P}_{c_k}^{NA,V}.$$

Since, $c_k \leq c$ implies $\mathcal{P}_{c_k}^{NA,V} \leq \mathcal{P}_c^{NA,V}$ and $\mathcal{D}_c^{NA,V} \geq \mathcal{D}_{c_k}^{NA,V}$ this allows us to deduce that

$$\mathcal{D}_c^{NA,V} \geq \mathcal{D}_{c_k}^{NA,V} = \mathcal{P}_{c_k}^{NA,V} \uparrow \mathcal{P}_c^{NA,V}. \qquad \square$$

In the proof of Proposition 5.5 we will use the min-max theorem in the following form.

Theorem 5.7 (see e.g. [55, Thm. 45.8] or [2, Thm. 2.4.1])**.** *Let K, L be convex subsets of vector spaces H_1 resp. H_2, where H_1 is locally convex and let $F : K \times L \to \mathbb{R}$ be given. If*

(1) K is compact,
(2) $F(\cdot, y)$ is continuous and convex on K for every $y \in L$,
(3) $F(x, \cdot)$ is concave on L for every $x \in K$

then

$$\sup_{y \in L} \inf_{x \in K} F(x, y) = \inf_{x \in K} \sup_{y \in L} F(x, y).$$

Proof of Proposition 5.5. We may assume that c is bounded from below by zero. Hence, by Lemma 5.6, it is sufficient to establish (5.12) for bounded continuous functions whose support satisfies

$$\mathsf{supp}\, c \subseteq C(\mathbb{R}_+) \times [0, t_0] \times \mathbb{R} \tag{5.13}$$

for some $t_0 \in \mathbb{R}_+$. Set

$$\mathsf{TM}_{t_0}^V(\mathbb{W}, \mu) := \{\pi \in \mathsf{TM}^V(\mathbb{W}, \mu) : \mathsf{supp}\, \pi \subseteq C(\mathbb{R}_+) \times [0, t_0] \times \mathbb{R}\},$$

$$\mathsf{DC}_c^{NA,V,t_0} := \left\{ (\varphi, \psi) \in C_b(\Omega) \times C_b(\mathbb{R}) : \exists \alpha \geq 0, \begin{array}{l} \varphi(\omega) + \psi(y) - \alpha(t - V) \leq c(\omega, t, y) \\ \text{for all } \omega \in \Omega, y \in \mathbb{R}, t \leq t_0 \end{array} \right\}.$$

Assume now that c satisfies (5.13). We then have

$$\inf_{\pi \in \mathsf{TM}^V(\mathbb{W},\mu)} \int c \, d\pi \;=\; \inf_{\pi \in \mathsf{TM}^{V,t_0}(\mathbb{W},\mu)} \int c \, d\pi, \tag{5.14}$$

$$\sup_{(\varphi,\psi) \in \mathsf{DC}_c^{NA,V}} \mathbb{W}(\varphi) + \mu(\psi) \;=\; \sup_{(\varphi,\psi) \in \mathsf{DC}_c^{NA,V,t_0}} \mathbb{W}(\varphi) + \mu(\psi). \tag{5.15}$$

In the following candidate sets we dispose of the moment condition on T and its dual counterpart:

$$\mathsf{TM}^{t_0}(\mathbb{W},\mu) := \{\pi : \mathsf{proj}_{C(\mathbb{R}_+)}(\pi) = \mathbb{W}, \mathsf{proj}_{\mathbb{R}}^i(\pi) = \mu, \mathrm{supp}\,\pi \subseteq C(\mathbb{R}_+) \times [0,t_0] \times \mathbb{R}\},$$

$$\mathsf{DC}_c^{NA,t_0} := \{(\varphi,\psi) \in C_b(\Omega) \times C_b(\mathbb{R}) : \varphi(\omega) + \psi(y) \le c(\omega,t,y) \text{ for } t \le t_0, y \in \mathbb{R}, \omega \in \Omega\}.$$

By the Monge-Kantorovich duality theorem in the form of Corollary 3.3 we have

$$\sup_{(\varphi,\psi) \in \mathsf{DC}_{\tilde{c}}^{NA,t_0}} \mathbb{W}(\varphi) + \mu(\psi) = \inf_{\pi \in \mathsf{TM}^{t_0}(\mathbb{W},\mu)} \int \tilde{c} \, d\pi \tag{5.16}$$

for $\tilde{c}$ lower semi-continuous and bounded from below. Using the min-max theorem (Theorem 5.7) with the function

$$F(\pi,\alpha) = \int c + \alpha(T - V) \, d\pi$$

for $\pi \in \mathsf{TM}^{t_0}(\mathbb{W},\mu)$ and $\alpha \ge 0$ we thus obtain

$$\inf_{\pi \in \mathsf{TM}^{V,t_0}(\mathbb{W},\mu)} \int c \, d\pi = \inf_{\pi \in \mathsf{TM}^{t_0}(\mathbb{W},\mu)} \int c \, d\pi + \sup_{\alpha \ge 0} \alpha \int T - V \, d\pi$$

$$= \sup_{\alpha \ge 0} \inf_{\pi \in \mathsf{TM}^{t_0}(\mathbb{W},\mu)} \int c + \alpha(T - V) \, d\pi \tag{5.17}$$

$$= \sup_{\alpha \ge 0} \sup_{(\varphi,\psi) \in \mathsf{DC}_{c+\alpha(T-V)}^{NA,t_0}} \mathbb{W}(\varphi) + \mu(\psi) \tag{5.18}$$

$$= \sup_{(\varphi,\psi) \in \mathsf{DC}_c^{NA,V,t_0}} \mathbb{W}(\varphi) + \mu(\psi),$$

where we have applied (5.16) to the function $\tilde{c} = c + \alpha(T - V)$ to establish the equality between (5.17) and (5.18). This concludes the proof. □

5.5. **Introducing Adaptedness.** Using the defining property of PRST, we are able to test the "adaptedness" of a probability π on $P(C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R})$ by integrating against martingales. For a continuous and bounded function $f : C(\mathbb{R}_+) \to \mathbb{R}_+$ we consider the S- continuous martingale f^M as in Proposition 4.11. Then $\pi \in \mathsf{TM}^V(\mathbb{W},\mu)$ satisfies $\pi \in \mathsf{JOIN}^{1,V}(\mathbb{W},\mu) = \mathsf{JOIN}^1(\mathbb{W},\mu) \cap \mathsf{TM}^V(\mathbb{W},\mu)$ if and only if for all continuous bounded functions $f : C(\mathbb{R}_+) \to \mathbb{R}, g : \mathbb{R} \to \mathbb{R}$

$$\int fg \, d\pi = \int f^M g \, d\pi. \tag{5.19}$$

This is a direct consequence of the definition of PRST and JOIN, see Definition 4.22 and Section 4.5.

Proposition 5.8. *Let $c : C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ be lower semi-continuous, predictable (cf. Remark 4.30) and bounded from below. Then*

$$\mathcal{D}_c^V = \mathcal{P}_c^V.$$

Proof. Using the same approximation procedure as in Lemma 5.6, we may assume that c is continuous and bounded. For the S- continuous martingale induced by a continuous and bounded function f we recall the notation $f^M = (f_t^M)_{t \in \mathbb{R}_+}$ introduced in Proposition 4.11. We want to use the min-max theorem, Theorem 5.7, with the function

$$F(\pi,h) = \int c + \bar{h} \, d\pi$$

where $\pi \in \mathsf{TM}^V(\mathbb{W},\mu)$ and

$$\bar{h}(\omega,t,y) = \sum_{i=1}^n (f_i(\omega) - f_i^M(\omega,t))g_i(y), \tag{5.20}$$

for $h(\omega, y) = \sum_{i=1}^{n} f_i(\omega)g_i(y)$, $n \in \mathbb{N}$, $f_i \in C_b(C(\mathbb{R}_+))$, $g_i \in C_b(\mathbb{R}_+)$.

The set $\mathsf{TM}^V(\mathbb{W}, \mu)$ is convex and compact by Prohorov's theorem and the set of all $\bar{h}$ of the form (5.20) is convex as well. Hence we obtain

$$
\begin{aligned}
\mathcal{P}_c^V &= \inf_{\pi \in \mathsf{JOIN}^{1,V}(\mathbb{W},\mu)} \int c \, d\pi \\
&= \inf_{\pi \in \mathsf{TM}^V(\mathbb{W},\mu)} \sup_h \int c + \bar{h} \, d\pi \\
&\overset{\text{Thm. 5.7}}{=} \sup_h \inf_{\pi \in \mathsf{TM}^V(\mathbb{W},\mu)} \int c + \bar{h} \, d\pi \\
&= \sup_h \sup_{(\varphi,\psi) \in \mathsf{DC}_{NA}^V(c+\bar{h})} \mathbb{W}(\varphi) + \mu(\psi),
\end{aligned}
$$

where the last equality holds by Proposition 5.5.

Write $c_h = c + \bar{h}$. For $(\varphi, \psi) \in \mathsf{DC}_{NA}^V(c_h)$ there is some $\alpha \geq 0$ such that

$$\varphi(\omega) + \psi(y) - \alpha(t - V) \leq c_h(\omega, t, y).$$

Taking conditional expectations w.r.t. $\mathcal{F}_t^0$ in the sense of Definition 4.10 we obtain

$$\varphi_t^M(\omega) + \psi(y) - \alpha(t - V) \leq c(\omega, t, y)$$

for all $\omega \in \Omega, t \in \mathbb{R}_+, y \in \mathbb{R}$ since c is predictable. This implies that $(\varphi^M, \psi) \in \mathsf{DC}_c^V$. Since $\varphi_0^M = \mathbb{W}(\varphi)$ we find

$$\mathcal{P}_c^V = \sup_h \sup_{(\varphi,\psi) \in \mathsf{DC}_{NA}^V(c+\bar{h})} (\mathbb{W}(\varphi) + \mu(\psi)) \leq \mathcal{D}_c^V. \tag{5.21}$$

As usual, the other inequality is straightforward. $\qquad\square$

Proof of Theorem 5.2. It is straight forward to see that c is lower semi-continuous since γ was assumed to be lower semi-continuous.

By Lemma 5.3 and Propositions 5.4 and 5.8 we have $\mathcal{D}_c^V \leq D_\gamma \leq P_\gamma \leq \mathcal{P}_c^V$ and $\mathcal{D}_c^V = \mathcal{P}_c^V$. Hence we have $D_\gamma = P_\gamma$ as required. $\qquad\square$

5.6. **General starting distribution.** In this section we consider $\tilde{\Omega} = \tilde{C}(\mathbb{R}_+)$, the set of all continuous functions on $\mathbb{R}_+$, and

$$\tilde{S} = \{(f, s) : f : [0, s] \to \mathbb{R} \text{ is continuous}\}.$$

Let λ be a probability measure on $\mathbb{R}$ prior to μ in convex order — i.e., $\int f(x) \, \lambda(dx) \leq \int f(x) \, \mu(dx)$ for any convex function $f(x)$. In particular λ is centered and $V_\lambda = \int x^2 \, \lambda(dx) \leq V < \infty$. Then there exist solutions to the Skorokhod embedding problem with general starting distribution and finite first moment. Denote by $\mathbb{W}_x$ the law of Brownian motion starting in x and put $\mathbb{W}_\lambda(d\omega) = \mathbb{W}_x(d\omega)\lambda(dx)$ for $\omega \in \tilde{\Omega}$, the law of Brownian motion starting at a random point according to the distribution λ. Given a functional $\gamma : \tilde{S} \to \mathbb{R}$ we are interested in the minimization problem

$$P_\gamma(\lambda, \mu) = \inf \left\{ \int \gamma(\omega, t)\xi(d\omega, dt), \xi \in \mathsf{RST}(\lambda, \mu) \right\}, \tag{5.22}$$

where $\mathsf{RST}(\lambda, \mu)$ is the set of all randomized stopping times ξ on $(\tilde{\Omega}, \mathbb{W}_\lambda)$ embedding μ and satisfying $\mathbb{E}_\xi[T] = V - V_\lambda$; in particular $\text{proj}_{\tilde{\Omega}}(\xi) = \mathbb{W}_\lambda$ and $h(\xi) = \mu$ for the map $h : \tilde{\Omega} \times \mathbb{R}_+ \to \mathbb{R}, (\omega, t) \mapsto \omega(t)$. We then have the following result:

Theorem 5.9. *Let $\gamma : \tilde{S} \to \mathbb{R}$ be $\tilde{S}$- lower semi-continuous and bounded from below in the sense of (5.2). Put*

$$D_\gamma(\lambda, \mu) = \sup \left\{ \int \psi(y) \, d\mu(y) : \psi \in C(\mathbb{R}), \begin{array}{l} \exists \varphi, \varphi \text{ is a } \tilde{S}\text{- continuous martingale,} \\ \mathbb{E}_{\mathbb{W}_\lambda}[\varphi_0] = 0, \varphi_t(\omega) + \psi(\omega(t)) \leq \gamma(\omega, t) \end{array} \right\}$$

where φ, ψ satisfy $|\varphi_t| \leq a + bt + cB_t^2$, $|\psi(y)| \leq a + by^2$ for some $a, b, c > 0$. Then we have the duality relation

$$P_\gamma(\lambda, \mu) = D_\gamma(\lambda, \mu). \tag{5.23}$$

The proof goes along the same lines as the proof of Theorem 5.2. The inequality $D_\gamma(\lambda, \mu) \leq P_\gamma(\lambda, \mu)$ is straightforward. For the other direction we can use the same argument as before. However, we have to replace $\mathbb{W}$ by $\mathbb{W}_\lambda$ and V by $\tilde{V} := V - V_\lambda$. Up to equation (5.9) everything can be copied line to line. Then we have to use the decomposition

$$\varphi_t(\omega) + \psi(\omega(t)) - \alpha(t - V + V_\lambda)$$
$$= [\varphi_t(\omega) + \alpha(\omega(t)^2 - t - V_\lambda)] + [\psi(\omega(t)) - \alpha(\omega(t)^2 - V)]$$

and note that $\mathbb{E}_{\mathbb{W}_\lambda}[\omega(t)^2] = t + V_\lambda$. The proof then concludes as before.

6. The monotonicity principle

In this section, we will establish Theorem 1.2. In fact we will show a stronger result that will imply Theorem 1.2. Moreover, in Theorem 6.8 and Section 6.3 we will prove further refinements of Theorem 1.2 that allow us to find sets $\Gamma \subseteq S$ with additional properties, for example recovering path properties of Brownian motion. We remark that the arguments used in this section are valid for arbitrary starting distributions.

The notion of stop-go pairs introduced in Definition 1.4 requires that *all* possible extensions (h, u) are considered. However, we want to consider a relaxed notion which is sensitive to the stopping measure ν. To this end we introduce the *conditional randomized stopping time given (f, s)*.

Definition 6.1. *Let $\xi \in \mathsf{RST}$ be given and* fix *a survival function H^ξ satisfying part (1) of Theorem 4.16. The conditional randomized stopping time of ξ, given $(f, s) \in S$, denoted by $\xi^{(f,s)}$, is defined to be*

$$\xi_\omega^{(f,s)}([0, t]) := \frac{1}{H^\xi(f, s)} \left(H^\xi(f, s) - H^\xi(f \oplus \omega_{\restriction [0, t]}, s + t) \right), \tag{6.1}$$

if $H^\xi(f, s) > 0$ and 1 otherwise.

This is the normalized stopping measure given that we followed the path f up to time s. In other words this is the normalized stopping measure of the "bush" which follows the "stub" (f, s). Note that we can equivalently write

$$\xi_\omega^{(f,s)}([0, t]) = \frac{1}{H^\xi(f, s)} \left(\xi_{f \oplus \omega}([0, t + s]) - \xi_{f \oplus \omega}([0, s]) \right).$$

Recall from (1.5) that for $\Gamma \subseteq S$ the set $\Gamma^<$ consists of all paths which have a proper extension that lies in Γ and that T denotes the projection from $C(\mathbb{R}_+) \times \mathbb{R}_+$ onto $\mathbb{R}_+$.

Lemma 6.2. *Let $\xi \in \mathsf{RST}^1$ be given and* fix *a survival function H^ξ satisfying part (1) of Theorem 4.16. Then, there is $\Gamma \subseteq S$ such that $r(\xi)(\Gamma) = 1$ and for all $(f, s) \in \Gamma^<$ we have $\xi^{(f,s)} \in \mathsf{RST}^1$. In addition, if $\mathbb{E}_\xi[T] < \infty$ then there is $\Gamma \subseteq S$ such that $\mathbb{E}_{\xi^{(f,s)}}[T] < \infty$ for all $(f, s) \in \Gamma^<$.*

We postpone the proof to a later point of this section and continue the discussion.

Definition 6.3. *Let $\xi \in \mathsf{RST}$ be given and* fix *a survival function H^ξ satisfying part (1) of Theorem 4.16. The set of* stop-go pairs relative to ξ *is defined by*

$$\mathsf{SG}^\xi = \Bigg\{ ((f, s), (g, t)) : f(s) = g(t), \tag{6.2}$$

$$\int \gamma(f \oplus \omega_{\restriction [0, r]}, s + r) \, d\xi^{(f,s)}(\omega, r) + \gamma(g, t) > \gamma(f, s) + \int \gamma(g \oplus \omega_{\restriction [0, r]}, t + r) \, d\xi^{(f,s)}(\omega, r) \Bigg\}.$$

The interpretation of SG^ξ is that in average it is better to stop at (f, s), chop off the "bush" and transfer it onto the "stub" (g, t).

Definition 6.4. *Let $\xi \in \mathsf{RST}^1$. Then a set $\Gamma \subseteq S$ is called γ-monotone iff*

$$\mathsf{SG}^\xi \cap (\Gamma^< \times \Gamma) = \emptyset.$$

Observe that a set is γ-monotone with respect to a particular stopping time. It will always be clear which stopping time is referred to from the context.

The following theorem implies Theorem 1.2 stated in the introduction.

Theorem 6.5. *Assume that $\gamma : S \rightarrow \mathbb{R}$ is Borel-measurable, the optimization problem (5.1) is well-posed and that $\nu \in \mathsf{RST}(\mu)$ is an optimizer. Then there exists a γ-monotone Borel set $\Gamma \subseteq S$ which supports ν in the sense that $r(\nu)(\Gamma) = 1$.*

The proof of this theorem relies on the following two results. The first result formalizes the heuristic idea that an optimizer cannot be improved on a large set of paths but at most on a small set of exceptional paths. The second result allows us to entirely exclude such an exceptional set of paths.

Recall the definition of $\mathsf{JOIN}(\mathbb{W}, \nu)$ from Section 4.5. We interpret the space $(C(\mathbb{R}_+) \times \mathbb{R}_+) \times (C(\mathbb{R}_+) \times \mathbb{R}_+)$ as a product $\mathcal{X} \times \mathcal{Y}$ so that we can make sense of the projections $\mathsf{proj}_\mathcal{X}$ and $\mathsf{proj}_\mathcal{Y}$. Note also that $(\mathcal{Y}, \nu) = (C(\mathbb{R}_+) \times \mathbb{R}_+, \nu)$ is a Polish probability space. An element $\pi \in \mathsf{JOIN}(\mathbb{W}, \nu)$ is a measure on $\mathcal{X} \times \mathcal{Y}$, and we will commonly want to consider the pushforward measure on $\mathcal{X}' \times \mathcal{Y}'$, given functions $f : \mathcal{X} \rightarrow \mathcal{X}'$, and $g : \mathcal{Y} \rightarrow \mathcal{Y}'$. We denote this measure by $(f \otimes g)(\pi)$. Typically f will be the map $r : C(\mathbb{R}_+) \times \mathbb{R}_+ \rightarrow S$, and g will be r or the identity.

Proposition 6.6. *Let ν be a randomized stopping time which minimizes (5.1) for a Borel measurable function $\gamma : S \rightarrow \mathbb{R}$. Assume that $\pi \in \mathsf{JOIN}(\mathbb{W}, \nu)$ satisfies*

$$H^{\mathsf{proj}_\mathcal{X}(\pi)}(f, s) > 0 \quad \Longrightarrow \quad H^\nu(f, s) > 0 \quad \text{for } (f, s) \in S. \tag{6.3}$$

Then we have $(r \otimes r)(\pi)(\mathsf{SG}^\nu) = 0$.

The interpretation of (6.3) is that if a particle has a strictly positive chance to be alive under $\mathsf{proj}_\mathcal{X}(\pi)$ then the probability that this particle is still alive under ν is positive as well.

Let τ be a non-randomized, bounded stopping time such that $[\![0, \tau]\!]$ is closed[6]. Then

$$\mathsf{M}_\tau := \{\xi \in \mathsf{M} : \xi(]\!]\tau, \infty]\!]) = 0\} \tag{6.4}$$

is compact as a consequence of Prohorov's theorem. We also let $\mathsf{RST}_\tau = \mathsf{RST} \cap \mathsf{M}_\tau$ and $\mathsf{PRST}_\tau = \mathsf{PRST} \cap \mathsf{M}_\tau$. The *joinings before* τ are the elements of the set

$$\mathsf{JOIN}(\tau, \nu) = \left\{ \pi \in \mathsf{JOIN}(\mathbb{W}, \nu) : \mathsf{proj}_{C(\mathbb{R}_+) \times \mathbb{R}_+}(\pi) \in \mathsf{PRST}_\tau \right\}. \tag{6.5}$$

For a non-randomized stopping time τ we write $\bar{\tau}$ for the random measure given through $\bar{\tau}(d\omega, dt) = \mathbb{W}(d\omega)\delta_{\tau(\omega)}(dt)$. We also recall from Definition 4.7 that a subset of S is left/right complete iff it is closed under forming restrictions/extensions of its paths.

Proposition 6.7. *Assume that τ is a (non-randomized) bounded $\mathcal{F}^+$-stopping time such that $[\![0, \tau]\!]$ is closed.*

Let $(\mathcal{Y}, \sigma)$ be a Polish probability space. Consider a set $B \subseteq S \times \mathcal{Y}$. If $(r \otimes \mathrm{Id})(\pi)(B) = 0$ for all $\pi \in \mathsf{JOIN}(\tau, \nu)$, then there exist a right complete set $D \subseteq S$ and a set $N \subseteq \mathcal{Y}$ such that $B \subseteq (D \times \mathcal{Y}) \cup (S \times N)$ and $r(\bar{\tau})(D) = \sigma(N) = 0$.

Proof of Theorem 6.5. Define a (non-randomized) stopping time τ_ν by

$$\tau_\nu(\omega) := \inf\{t : H^\gamma \circ r(\omega, t) = 0\}.$$

Using Lemma 4.13 and Corollary 4.15, by choosing a suitable foretelling sequence of the stopping times $\inf\{t : H^\nu(r(\omega, t)) \le 1/n\}$, we can pick a sequence $\tau_n, n \ge 1$ of $\mathcal{F}^+$-stopping times such that

(1) $\tau_n \uparrow \tau_\nu$.
(2) $\tau_n \le n$.
(3) $[\![0, \tau_n]\!]$ is closed.
(4) $\tau_n < \inf\{t : H^\nu(r(\omega, t)) \le 1/n\}$ on $\{\omega : H^\nu(r(\omega, 0)) > 1/n\}$.

---

[6]We emphasize that this means that $[\![0, \tau]\!]$ is closed as a subset of $C(\mathbb{R}_+) \times \mathbb{R}_+$.

Fix n. Then every joining $\pi \in \mathsf{JOIN}(\tau_n, \nu)$ satisfies the assumptions of Proposition 6.6 and hence $(r \otimes r)(\pi)(\mathsf{SG}^\nu) = 0$ for every $\pi \in \mathsf{JOIN}(\tau_n, \nu)$. We specify $(\mathcal{Y}, \sigma) = (S, r(\nu))$ in Proposition 6.7 and apply it to the stopping times $(\tau_n)_{n \in \mathbb{N}}$ to find left complete sets $L_n := S \setminus D_n$ and sets $\Gamma_n := S \setminus N_n$ such that

$$\mathsf{SG}^\nu \cap (L_n \times \Gamma_n) = \emptyset$$

and $r(\bar{\tau}_n)(L_n) = 1, r(\nu)(\Gamma_n) = 1$. Setting $\tilde{\Gamma} := \bigcap_n \Gamma_n$ we have $\mathsf{SG}^\nu \cap (L_n \times \tilde{\Gamma}) = \emptyset$ for all n. With $L := \bigcup_n L_n$ we have

$$\mathsf{SG}^\nu \cap (L \times \tilde{\Gamma}) = \emptyset. \tag{6.6}$$

Put

$$L^+ := \{(g, t) \; : \; (g_{\upharpoonright [0,s]}, s) \in L \text{ for all } s < t\}.$$

Note that L^+ is universally measurable: its complement $S \setminus L^+$ is given by

$$\{(g, t) : \exists s < t, (g_{\upharpoonright [0,s]}, s) \notin L\} \tag{6.7}$$

$$= \mathrm{proj}_S \{((g, t), s) \in S \times [0, t) : (g_{\upharpoonright [0,s]}, s) \notin L\} \tag{6.8}$$

and is hence analytic.

Because L is left complete we have $L \subseteq L^+$ and also $(L^+)^< \subseteq L$. Clearly $r(\bar{\tau}_n)(L) = r(\bar{\tau}_n)(L^+) = 1$ for all n. Moreover, the set $\{t : r(\omega, t) \in L^+\}$ is either empty or closed. Hence, the convergence $\tau_n \nearrow \tau_\nu$ almost surely implies that $r(\bar{\tau}_\nu)(L^+) = 1$ and therefore $r(\nu)(L^+) = 1$. Hence, setting $\bar{\Gamma} := L^+ \cap \tilde{\Gamma}$ we have $r(\nu)(\bar{\Gamma}) = 1$, $\bar{\Gamma}^< \subseteq (L^+)^< \subseteq L$ and by (6.6) we can conclude

$$\mathsf{SG}^\nu \cap (\bar{\Gamma}^< \times \bar{\Gamma}) = \emptyset.$$

Of course this pertains if we replace $\bar{\Gamma}$ by a Borel subset of full measure.   □

With a slightly easier proof we get the following important sibling of Theorem 6.5.

Theorem 6.8. *Let $\xi \in \mathsf{RST}^1$ and fix a survival function H^ξ satisfying part (1) of Theorem 4.16. Assume that $U \subseteq S$ is such that $r(\pi)(U) = 0$ for all $\pi \in \mathsf{RST}$ satisfying*

$$r(\pi)(\{(f, s) \in S \; : \; H^\xi(f, s) = 0\}) = 0. \tag{6.9}$$

Then, there is a set $\Gamma \subseteq S$ such that $r(\xi)(\Gamma) = 1$ and $U \cap \Gamma^< = \emptyset$.

Proof. The argument goes along the lines of the previous proof, considering the Polish space $\mathcal{Y} = \{y\}$ equipped with the probability measure $\sigma = \delta_y$.   □

Proof of Lemma 6.2. We want to apply the previous theorem. Consider

$$U_1 = \left\{(f, s) \in S \; : \; H^\xi(f, s) > 0, \int d\xi^{(f,s)}(\omega, t) < 1\right\},$$

$$U_2 = \left\{(f, s) \in S \; : \; H^\xi(f, s) > 0, \int t \, d\xi^{(f,s)}(\omega, t) = \infty\right\}.$$

We have to show that $\pi(U_1 \cup U_2) = 0$ for all $\pi \in \mathsf{RST}$ satisfying (6.9). Fix such a π. We first show that $\pi(U_1) = 0$, by contradiction. Assume that $\pi(U_1) > 0$. Let ρ be the representation of π given in Theorem 4.16 (4) and α the representation of ξ. Then there exists a set $A \in \bar{\mathcal{F}}_\rho$ with $\bar{\mathbb{P}}(A) > 0$ such that

$$A_S := \{r(\omega, \rho(\omega, u)) : (\omega, u) \in A\} \subseteq U_1.$$

Define $M_t := \bar{\mathbb{E}}[\mathbb{1}_A | \bar{\mathcal{F}}_t]$. Then, $(M_{\rho+s})_{s \geq 0}$ is an $\bar{\mathcal{F}}_{\rho+s}$ martingale as $A \in \bar{\mathcal{F}}_\rho$. Set $\kappa := \rho \vee \alpha$. Applying optional stopping twice and using the fact that κ is almost surely finite, we get $\bar{\mathbb{E}} M_\kappa = \bar{\mathbb{E}} M_\rho$. However,

$$\bar{\mathbb{E}}(M_\kappa - M_\rho) = \int_{A_S} H^\xi(f, s) \left(\int d\xi^{(f,s)}(\omega, t) - 1\right) d\pi((f, s)) < 0,$$

by the choice of A and κ and Assumption (6.9). This is a contradiction.

Next, we assume that $\mathbb{E}_\xi[T] < \infty$ and that there is $\pi \in \mathsf{RST}$ satisfying (6.9) such that $\pi(U_2) > 0$. Then, we can compute using the definition of U_2

$$\infty > \mathbb{E}_\xi[T] \geq \int_{U_2} H^\xi(f, s) \int t \, d\xi^{(f,s)}(\omega, t) \, d\pi(f, s) = \infty,$$

which is a contradiction. □

6.1. **Proof of Proposition 6.6.**

Proof of Proposition 6.6. Note that if $v' \in \mathsf{RST}(\mu')$ and $v'' \in \mathsf{RST}(\mu'')$, then $(v' + v'')/2 \in \mathsf{RST}((\mu' + \mu'')/2)$. The probabilistic interpretation of this is that the random stopping time $(v' + v'')/2$ corresponds to flipping a coin at time $t = 0$ and subsequently either applying the randomized stopping rule v' or v''.

Working towards a contradiction we assume that there is $\pi \in \mathsf{JOIN}(\mathbb{W}, v)$ such that $(r \otimes r)(\pi)(\mathsf{SG}^v) > 0$. By looking at $(r \otimes r)(\bar{\pi}) := (r \otimes r)(\pi)_{|\mathsf{SG}^v}$ we can assume that $(r \otimes r)(\pi)$ is concentrated on SG^v. As SG^v corresponds to a predictable subset of $\mathcal{X} \times \mathcal{Y} = (C(\mathbb{R}_+) \times \mathbb{R}_+) \times \mathcal{Y}$ (recall Proposition 4.4 and Remark 4.30) we can assume that $\mathsf{proj}_{\mathcal{X}}(\pi) \in \mathsf{RST}$. As $H^v(f, s) = 0$ implies $v^{(f,s)}(A) = \mathbb{W}(\mathsf{proj}_\Omega(A \cap \{\Omega \times \{s\}\}))$ for any Borel set $A \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+$ (cf. Definition 6.1), i.e. there is instant stopping, there is no stop-go pair $((f, s), (g, t))$ with $H^v(f, s) = 0$. Hence, $r(\mathsf{proj}_{\mathcal{X}}(\pi))(\{(f, s) : H^v(f, s) = 0\}) = 0$.

Set $v_0 = v_1 := v$. We then use π to define two modifications v_0^π and v_1^π of v such that the following hold true:

(1) The terminal distributions μ_0, μ_1 corresponding to v_0^π and v_1^π satisfy $(\mu_0 + \mu_1)/2 = \mu$.
(2) v_0^π stops paths earlier than $v_0 = v$ while v_1^π stops later than $v_1 = v$.
(3) The cost of v_0^π plus the cost of v_1^π is less than twice the cost of v, i.e.

$$\int \gamma(\omega_{|[0,t]}, t)\, dv_0^\pi(\omega, t) + \int \gamma(\omega_{|[0,t]}, t)\, dv_1^\pi(\omega, t) < 2 \int \gamma(\omega_{|[0,t]}, t)\, dv(\omega, t).$$

More formally, (2) asserts that for every $s \geq 0$,

$$(v_0^\pi)_\omega[0, s] \geq v_\omega[0, s], \text{ a.s.} \tag{6.10}$$

$$\text{and } (v_1^\pi)_\omega[0, s] \leq v_\omega[0, s], \text{ a.s.,} \tag{6.11}$$

where $v_{\omega \in \Omega}, (v_0^\pi)_{\omega \in \Omega}, (v_1^\pi)_{\omega \in \Omega}$ are disintegrations of v_0, v_0^π, v_1^π respectively w.r.t. $\mathbb{W}$.

If we are able to construct such a pair v_0^π, v_1^π, then $(v_0^\pi + v_1^\pi)/2$ is a randomized stopping time in $\mathsf{RST}(\mu)$ which is strictly better than v and therefore yields the desired contradiction.

To define v_0^π, we first consider $\rho_0 = \mathsf{proj}_{\mathcal{X}}(\pi)$ which is a randomized stopping time. As in Remark 4.18 we can view ρ_0 as a right-continuous decreasing survival function $H^{\rho_0} : S \to [0, 1]$ which starts at 1. It is possible that H^{ρ_0} does not decrease to 0 since we allow particles to survive until ∞.

We now define the randomized stopping time v_0^π as the product

$$H^{v_0^\pi}(f, s) := H^{\rho_0}(f, s) \cdot H^v(f, s).$$

The probabilistic interpretation of this definition is that a particle is stopped by v_0^π if it is stopped by ρ_0 or stopped by v, where these events are taken to be conditionally independent given the particle followed the path f until time s. Comparing v_0 and v_0^π the latter will stop some particles earlier than the first one. We note that this in particular implies that $\mathbb{E}_{v_0^\pi}[T] \leq \mathbb{E}_v[T] < \infty$, where T is the projection from $C(\mathbb{R}_+) \times \mathbb{R}_+$ onto $\mathbb{R}_+$ as defined in (4.7). Also clearly, $v_0^\pi \in \mathsf{RST}$, i.e. v_0^π inherits adaptivity from ρ_0 and v. Equivalently we can define v_0^π by setting for $A \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+$

$$v_0^\pi(A) = \int_A H^v(\omega, t)\, d\rho_0(\omega, t) + \int_A H^{\rho_0}(\omega, t)\, dv(\omega, t).$$

Let us now turn to the definition of v_1^π. For $A \subseteq C(\mathbb{R}_+) \times \mathbb{R}_+$ we define

$$\rho_1(A) = \int_{\mathcal{X} \times A} H^v(\omega, s)\, d\pi((\omega, s), (\eta, t)).$$

Fix an $\mathcal{F}^0$-measurable disintegration $(v_\omega)_{\omega \in C(\mathbb{R}_+)}$ of v by (1) of Theorem 4.16. Given $(f, s) \in S$ and $(\omega, t) \in C(\mathbb{R}_+) \times \mathbb{R}_+$ we define a measure on $\mathbb{R}$ with support in $[t, \infty)$ by setting for $I \subseteq [t, \infty)$

$$v_{(f,s),(\omega,t)}(I) := v_{f \oplus \theta_t(\omega)}(I - t + s) = H^v(f, s)\, v_{\theta_t(\omega)}^{(f,s)}(I - t + s), \tag{6.12}$$

where $\theta_t(\omega) = (\omega_{s+t} - \omega_t)_{s\geq 0}$. Note that this is a slight generalization of a conditional randomized stopping time, see (6.1). Here we additionally allow a shift of the time parameter and do not normalize (hence the additional factor $H^\nu(f, s)$). This is necessary as in the next step – for defining ν_1^π – we need to trim bushes; i.e. we need to cut some paths at time s and plant them on a stub at time t. Additionally, we can only move the mass that is present which accounts for the $H^\nu(f, s)$ appearing in (6.12) and the definition of ρ_1. Moreover, note that for a set $I \subseteq (t, t + u)$ given (f, s) the map

$$(\omega, t) \mapsto \nu_{(f,s),(\omega,t)}(I)$$

is $\sigma(\omega_l, t \leq l \leq t + u)$-measurable.

We define the probability measure ν_1^π on $C(\mathbb{R}_+) \times \mathbb{R}_+$ by

$$\nu_1^\pi(A) = \nu_1(A) - \rho_1(A) + \int_{X \times A} \nu_{(\omega_{\restriction[0,s]},s),(\eta,t)}(A_\eta) \, d\pi((\omega, s), (\eta, t)),$$

where $A_\eta = \{t \in \mathbb{R}_+ : (\eta, t) \in A\}$. The interpretation of this definition is the following. The support of the randomized stopping time ν can be thought as a tree. The joining π defines a plan how to trim the tree, i.e. cut a bush at position (f, s) and plant it on top of $(\eta_{\restriction[0,t]}, t)$. Hence, we take the tree, ν, prepare the position where something will be newly planted, subtract ρ_1 which takes away some mass, and plant as much as possible on these stubs to end up with a tree of mass one again. Due to the measurability properties of $(\omega, t) \mapsto \nu_{(\omega,t)}^{(f,s)}$ we obtain that $\nu_1^\pi \in \mathsf{RST}$. Moreover, as $\mathrm{proj}_\mathcal{Y}(\pi) \leq \nu$ by the definition of $\mathsf{JOIN}(\mathbb{W}, \nu)$ we get $\mathbb{E}_{\nu_1^\pi}[T] \leq 2\mathbb{E}_\nu[T] < \infty$.

Summing up we have constructed ν_0^π, $\nu_1^\pi \in \mathsf{RST}$ such that $\mathbb{E}_{\nu_0^\pi}[T], \mathbb{E}_{\nu_1^\pi}[T] < \infty$. It remains to show that $\nu^\pi = 1/2(\nu_0^\pi + \nu_1^\pi) \in \mathsf{RST}(\mu)$ and that $\int \gamma \, d\nu^\pi < \int \gamma \, d\nu$. To this end let us consider the constrictions of ν_0^π and ν_1^π separately. For $A \subseteq \Omega \times \mathbb{R}_+$ we have

$$\nu_0^\pi(A) - \nu(A) = \int_A H^\nu(\omega, t) \, d\rho_0(\omega, t) - \int_A (\rho_0)_\omega([0, t]) \, d\nu(\omega, t).$$

Furthermore,

$$\int_A (\rho_0)_\omega([0, t]) \, d\nu(\omega, t)$$
$$= \int_A \int_\mathcal{Y} \int_0^t d\pi((\omega, u), (\eta, s)) \, d\nu_\omega(t) \, d\mathbb{W}(\omega)$$
$$= \int_\Omega \int_\mathcal{Y} \int_{\mathbb{R}_+} \int_u^\infty \mathbb{1}_A(\omega, t) \, d\nu_\omega(t) \, d\pi((\omega, u), (\eta, s)) \, d\mathbb{W}(\omega)$$
$$= \int_\Omega \int_\mathcal{Y} \int_{\mathbb{R}_+} H^\nu(\omega, u) \nu_{\theta_u(\omega)}^{(\omega_{\restriction[0,u]},u)}(A_\omega) \, d\pi((\omega, u), (\eta, s)) \, d\mathbb{W}(\omega).$$

This yields

$$\int \gamma \, d(\nu_0^\pi - \nu) \tag{6.13}$$
$$= \int d\pi((\omega, s), (\eta, t)) \, H^\nu(\omega, s) \left[\gamma(\omega_{\restriction[0,s]}, s) - \int \gamma(\omega_{\restriction[0,s]} \oplus \tilde{\omega}_{\restriction[0,u]}, s + u) \, d\nu^{(\omega_{\restriction[0,s]},s)}(\tilde{\omega}, u) \right].$$

For ν_1^π we can compute

$$\int \gamma \, d(\nu_1^\pi - \nu) \tag{6.14}$$
$$= \int d\pi((\omega, s), (\eta, t)) \, H^\nu(\omega, s) \left[\int \gamma(\eta_{\restriction[0,t]} \oplus \tilde{\omega}_{\restriction[0,u]}, t + u) \, d\nu^{(\omega_{\restriction[0,s]},s)}(\tilde{\omega}, u) - \gamma(\eta_{\restriction[0,t]}, t) \right].$$

We thus obtain

$$2 \int \gamma \, d(\nu^\pi - \nu)$$

$$= \int d\pi((\omega, s), (\eta, t)) \Big(H^\nu(\omega, s) \Big[- \int \gamma(\omega_{\restriction[0,s]} \oplus \tilde{\omega}_{\restriction[0,r]}, s + r) \, d\nu^{(\omega_{\restriction[0,s]}, s)}(\tilde{\omega}, r) - \gamma(\eta_{\restriction[0,t]}, t) +$$

$$\gamma(\omega_{\restriction[0,s]}, s) + \int \gamma(\eta_{\restriction[0,t]} \oplus \tilde{\omega}_{\restriction[0,r]}, t + r) \, d\nu^{(\omega_{\restriction[0,s]}, s)}(\tilde{\omega}, r) \Big] \Big),$$

which is strictly negative by the definition of stop-go pairs relative to ν and Assumption (6.3). Moreover, the last identity holds for arbitrary bounded $F : C(\mathbb{R}_+) \times \mathbb{R}_+ \to \mathbb{R}$ instead of γ. In particular, taking $F(\omega, t) = G(\omega(t))$ for $G : \mathbb{R} \to \mathbb{R}$ bounded we get

$$2 \int G \, d(\nu^\pi - \nu)$$

$$= \int d\pi((\omega, s), (\eta, t)) \Big(H^\nu(\omega, s) \Big[- \int G(\omega_{\restriction[0,s]} \oplus \tilde{\omega}_{\restriction[0,r]}(s + r)) \, d\nu^{(\omega_{\restriction[0,s]}, s)}(\tilde{\omega}, r) - G(\eta(t)) +$$

$$G(\omega(s)) + \int G(\eta_{\restriction[0,t]} \oplus \tilde{\omega}_{\restriction[0,r]}(t + r)) \, d\nu^{(\omega_{\restriction[0,s]}, s)}(\tilde{\omega}, r) \Big] \Big) = 0,$$

because $((f, s), (g, t)) \in \mathsf{SG}^\nu$ implies that $f(s) = g(t)$ and $(r \otimes r)(\pi)$ is concentrated on SG^ν. This proves $\nu^\pi \in \mathsf{RST}(\mu)$. Hence, we obtain the desired contradiction. $\qquad \square$

6.2. **Proof of Proposition 6.7.**

Important Convention. For the remainder of this section we fix a (finite) non-randomized stopping time τ such that $[\![0, \tau]\!]$ is closed and satisfies $\tau \le t_0$ for some $t_0 \in \mathbb{R}_+$.

6.2.1. *An auxiliary Optimization Problem.*

We fix a Polish probability space $(\mathcal{Y}, \sigma)$. Let $c : C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y} \to \mathbb{R}_+$ be a predictable upper semi-continuous function. We are interested in the maximization problem

$$P^{\le 1} = P_c^{\le 1}(\mathbb{W}, \tau, \sigma) = \sup_{\pi \in \mathsf{JOIN}(\tau, \sigma)} \int_{C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y}} c \, d\pi \tag{6.15}$$

and its relation to the dual problem

$$D^{\le 1} = D_c^{\le 1}(\mathbb{W}, \tau, \sigma) = \inf_{(\varphi, \psi) \in \mathsf{DC}} \Big(\mathbb{W}(\varphi_\tau^M) + \sigma(\psi) \Big), \tag{6.16}$$

$$\mathsf{DC} = \{(\varphi, \psi) \in C_b(\Omega) \times C_b(\mathcal{Y}) : \varphi, \psi \ge 0, c(\omega, t, y) \le \varphi_t^M(\omega) + \psi(y), t \le \tau, y \in \mathcal{Y}, \omega \in \Omega\},$$

where φ^M is an S- continuous martingale (cf. Proposition 4.11). To indicate the dependence of DC on the cost function c we sometimes write $\mathsf{DC}(c)$. Note that for integrable φ we always have $\mathbb{W}(\varphi) = \mathbb{W}(\varphi_\tau^M)$ by optional stopping.

Observe that due to predictability of c the maximization problem is not altered if we replace PRST_τ by RST_τ in the definition of $\mathsf{JOIN}(\tau, \sigma)$, cf. (6.5).

As above (Lemma 5.3), the inequality $D^{\le 1} \ge P^{\le 1}$ is trivial. We now consider the other inequality.

Proposition 6.9. *Let $c : C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y} \to \mathbb{R}_+$ be predictable (in the sense of Remark 4.30), upper semi-continuous and bounded from above. Assume that τ is a bounded stopping time such that $[\![0, \tau]\!]$ is closed. Then*

$$P^{\le 1} = \sup_{\pi \in \mathsf{JOIN}(\tau, \sigma)} \int_{C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y}} c \, d\pi = \inf_{(\varphi, \psi) \in \mathsf{DC}} (\mathbb{W}(\varphi) + \sigma(\psi)) = D^{\le 1}.$$

We first establish a variant which applies to not necessarily predictable c. Then, we will use the defining property of PRST, Equation (4.6), to derive the predictable version.

Proposition 6.10. *Let* $c : C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y} \to \mathbb{R}_+$ *be (upper semi-) continuous and bounded from above. Then*

$$P^{\leq 1, NA} := \sup_{\pi \in \mathsf{TM}(\tau, \sigma)} \int_{C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y}} c \, d\pi = \inf_{(\varphi, \psi) \in \widetilde{\mathsf{DC}}} (\mathbb{W}(\varphi) + \sigma(\psi)) =: D^{\leq 1, NA},$$

where $\widetilde{\mathsf{DC}} = \{(\varphi, \psi) \in C_b(\Omega) \times C_b(\mathcal{Y}) : \varphi, \psi \geq 0, c(\omega, t, y) \leq \varphi(\omega) + \psi(y) \text{ for all } y \in \mathcal{Y}, t \leq \tau, \omega \in \Omega\}$.

Here the set of all *tagged random measures* is given by

$$\mathsf{TM}(\tau, \sigma) := \Big\{ \pi \in \mathcal{P}^{\leq 1}(C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y}) : \mathsf{proj}_{C(\mathbb{R}_+) \times \mathbb{R}_+}(\pi) \in \mathsf{M}_\tau, \mathsf{proj}_{\mathcal{Y}}(\pi) \leq \sigma \Big\}.$$

Proof of Proposition 6.10. We reduce the theorem to the usual transport duality result. Put $\bar{c}(\omega, y) = \sup_{t \leq \tau(\omega)} c(\omega, t, y)$. As $[\![0, \tau]\!]$ is closed and bounded $\bar{c}$ is continuous.

The dual constraint set can be written as

$$\widetilde{\mathsf{DC}} = \{(\varphi, \psi) \in C_b(\Omega) \times C_b(\mathcal{Y}) : \varphi, \psi \geq 0, \bar{c}(\omega, y) \leq \varphi(\omega) + \psi(y) \text{ for all } y \in \mathcal{Y}, \omega \in \Omega\}.$$

From the classical duality theorem of optimal transport (3.1) we know that

$$\inf_{(\varphi, \psi) \in \widetilde{\mathsf{DC}}} \mathbb{W}(\varphi) + \sigma(\psi) = \sup_{q \in \mathsf{Cpl}(\mathbb{W}, \sigma)} \int_{\Omega \times \mathcal{Y}} \bar{c}(\omega, y) \, q(d\omega, dy) =: \check{P}.$$

It remains to show that $\check{P} = P^{\leq 1, NA}$. From the definition of $\bar{c}$ and $\mathsf{TM}(\tau, \sigma)$ it is clear that we always have $P^{\leq 1, NA} \leq \check{P}$. To prove the other inequality fix $\varepsilon > 0$ and take $q \in \mathsf{Cpl}(\mathbb{W}, \sigma)$. For any (ω, y) there is $t(\omega, y) \leq \tau(\omega)$ such that $c(\omega, t(\omega, y), y) + \varepsilon \geq \bar{c}(\omega, y)$ and we may assume that t depends measurably on (ω, y). Putting $\pi(d\omega, ds, dy) := q(d\omega, dy)\delta_{t(\omega, y)}(ds) \in \mathsf{TM}$ we get

$$\int_{C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathcal{Y}} c(\omega, s, y) \, \pi(d\omega, ds, dy) + \varepsilon \geq \int_{\Omega \times \mathcal{Y}} \bar{c}(\omega, y) \, q(d\omega, dy).$$

This implies that $P^{\leq 1, NA} + \varepsilon \geq \check{P}$. Letting ε go to zero we obtain the claim. $\qquad\square$

Proof of Proposition 6.9. As c is bounded from above we have $P^{\leq 1} < \infty$. Arguing as in Lemma 5.6, we may assume that the cost function c is continuous.

We will now proceed as in the proof of Proposition 5.8, i.e. we consider again the functions $h, \bar{h}$ as in (5.20) and we shall apply Theorem 5.7 to the function

$$F(\pi, h) = \int c + \bar{h} \, d\pi$$

for $\pi \in \mathsf{TM}(\tau, \sigma)$. The set $\mathsf{TM}(\tau, \sigma)$ is convex and compact by Prohorov's theorem and the set of h under consideration is convex as well. The function F is continuous and affine in π and h. This allows us to deduce

$$P^{\leq 1} = \sup_{\pi \in \mathsf{JOIN}(\tau, \sigma)} \int c \, d\pi$$

$$= \sup_{\pi \in \mathsf{TM}(\tau, \sigma)} \inf_h \int c + \bar{h} \, d\pi$$

$$\stackrel{\text{Thm. 5.7}}{=} \inf_h \sup_{\pi \in \mathsf{TM}(\tau, \sigma)} \int c + \bar{h} \, d\pi$$

$$= \inf_h \inf_{(\varphi, \psi) \in \widetilde{\mathsf{DC}}(c + \bar{h})} [\mathbb{W}(\varphi) + \sigma(\psi)],$$

where the last equality holds by Proposition 6.10. Write

$$c_h(\omega, t, y) = c(\omega, t, y) + \sum_{i=1}^n (f_i(\omega) - f_{i,t}^M(\omega))g(y).$$

For $(\varphi, \psi) \in \widetilde{\mathsf{DC}}(c_h)$ we have

$$c_h(\omega, t, y) \leq \varphi(\omega) + \psi(y).$$

Taking conditional expectations w.r.t. $\mathcal{F}_t^0$ in the sense of Definition 4.10 and using predictability of c we get

$$c(\omega, t, y) \le \varphi_t^M(\omega) + \psi(y),$$

i.e. $(\varphi, \psi) \in \mathsf{DC}(c)$, hence $\widetilde{\mathsf{DC}}(c_h) \subseteq \mathsf{DC}(c)$. Because $\mathbb{W}(\varphi_t^M) = \mathbb{W}(\varphi)$ this implies that

$$P^{\le 1} = \inf_h \inf_{(\varphi,\psi)\in\widetilde{\mathsf{DC}}(c+\bar{h})} \left[\mathbb{W}(\varphi) + \sigma(\psi)\right]$$

$$\ge \inf_{(\varphi,\psi)\in\mathsf{DC}(c)} \left[\mathbb{W}(\varphi) + \sigma(\psi)\right] = D^{\le 1}.$$

Trivially $D^{\le 1} \ge P^{\le 1}$, hence we conclude $D^{\le 1} = P^{\le 1}$. $\qquad\square$

6.2.2. *A Choquet argument.* Denote by $\mathsf{LSC}_b(X)$ the set of bounded lower semi-continuous functions on X. The following lemma is a simple consequence of Proposition 6.9.

Lemma 6.11. *Let* $c : C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y} \to [0, 1]$ *be predictable (in the sense of Remark 4.30) and upper semi-continuous. Assume that τ is a bounded stopping time such that $[\![0,\tau]\!]$ is closed. Then*

$$P^{\le 1} = \sup_{\pi\in\mathsf{JOIN}(\tau,\sigma)} \int_{C(\mathbb{R}_+)\times\mathbb{R}_+\times\mathcal{Y}} c\, d\pi = \inf_{(\varphi,\psi)\in\mathsf{DC}'} (\mathbb{W}(\varphi) + \sigma(\psi)) = D',$$

where DC' *is given by*

$$\left\{(\varphi,\psi) \in \mathsf{LSC}(\Omega) \times \mathsf{LSC}(\mathcal{Y}) : 0 \le \varphi, \psi \le 1, c(\omega, t, y) \le \varphi_t^M(\omega) + \psi(y), t \le \tau, y \in \mathcal{Y}, \omega \in \Omega\right\}.$$

Proof. By Proposition 6.9, pick two continuous, bounded and non-negative functions $\varphi, \psi \in \mathsf{DC}$. Then $\bar{\psi} := \psi \wedge 1$ is still continuous and as $\varphi \ge 0$ we also have $(\varphi, \bar{\psi}) \in \mathsf{DC}$. Put $\rho = \inf\{t \ge 0 : \varphi_t^M > 1\}$. Due to continuity of φ^M the set $D := \{\varphi^M > 1\}$ is open. Hence also $\{\rho < \infty\} = \mathrm{proj}_\Omega D$ is open as projections are open mappings. The map $\omega \mapsto \varphi_{\rho(\omega)}^M(\omega) =: \bar{\varphi}(\omega) \le 1$ is lower semi-continuous. Clearly, $(\bar{\varphi}, \bar{\psi}) \in \mathsf{DC}'$ with $\mathbb{W}(\bar{\varphi}) + \sigma(\bar{\psi}) \le \mathbb{W}(\varphi) + \sigma(\psi)$. $\qquad\square$

As the indicator function of a closed set is upper semi-continuous Lemma 6.11 implies

Corollary 6.12. *Let $K \subseteq S \times \mathcal{Y}$ be closed. Assume that τ is a bounded stopping time such that $[\![0,\tau]\!]$ is closed. Then*

$$\sup_{\pi\in\mathsf{JOIN}(\tau,\sigma)} \int_{S\times\mathcal{Y}} \mathbb{1}_K \circ (r \otimes \mathrm{Id})\, d\pi = \inf_{(\varphi,\psi)\in\mathsf{DC}''} \left[\mathbb{W}(\varphi) + \sigma(\psi)\right],$$

where DC'' consists of all pairs $(\varphi, \psi) \in \mathsf{LSC}(\Omega) \times \mathsf{LSC}(\mathcal{Y})$ satisfying

$$0 \le \varphi, \psi \le 1, \mathbb{1}_K((f, s), y) \le \varphi^M(f, s) + \psi(y), (f, s) \in r([\![0,\tau]\!]), y \in \mathcal{Y}.$$

Through an application of Choquet's theorem we will extend this to

Lemma 6.13. *Let $K \subseteq S \times \mathcal{Y}$ be a Borel set. Assume that τ is a bounded stopping time such that $[\![0,\tau]\!]$ is closed. Then*

$$P^{\le 1} = \sup_{\pi\in\mathsf{JOIN}(\tau,\sigma)} \int_{S\times\mathcal{Y}} \mathbb{1}_K \circ (r \otimes \mathrm{Id})\, d\pi = \inf_{(\varphi,\psi)\in\mathsf{DC}''} (\mathbb{W}(\varphi) + \sigma(\psi)) = D''. \qquad (6.17)$$

Proof. To indicate the dependence on the set K we write $D''(K)$ and for notational convenience we drop the two primes and simply write $D(K)$. The left hand side in (6.17) is clearly a capacity on $S \times \mathcal{Y}$. To establish the claim, it is therefore sufficient to show that $D(K)$ is also a capacity on $S \times \mathcal{Y}$, because the indicator of a closed set is upper semi-continuous and the result then follows from Corollary 6.12 and Choquet's theorem.

Hence, we need to show the three defining properties of a capacity, namely monotonicity, continuity from below and continuity from above for compact sets. The monotonicity is clear. Let us turn to the continuity from below.

Take an increasing sequence $A_1 \subseteq A_2 \subseteq \ldots \subseteq S \times \mathcal{Y}$ of Borel sets and put $A = \bigcup_n A_n$. For all n there are lower semi-continuous functions[7] $\varphi_n : C(\mathbb{R}_+) \to [0, 1]$ and $\psi_n : \mathcal{Y} \to [0, 1]$ such that $\mathbb{1}_{A_n}((f, s), y) \leq \varphi_n^M(f, s) + \psi_n(y)$ for all $(f, s) \in S, y \in \mathcal{Y}$ and

$$\sigma(\psi_n) + \mathbb{W}(\varphi_n) \leq D(A_n) + 1/n.$$

Using a Mazur/Komlos-type lemma we can assume that some appropriate convex combinations of ψ_n and φ_n converge a.s. to functions ψ and φ. Let us be a little bit more precise here. There exist convex coefficients $\alpha_n^n, \ldots, \alpha_{k_n}^n, n \geq 1, k_n < \infty$, and full measure subsets $\Omega_1 \subseteq \Omega, \mathcal{Y}_1 \subseteq \mathcal{Y}$ such that with $\tilde{\varphi}_n := \sum_{i=n}^{k_n} \alpha_i^n \varphi_i, \tilde{\psi}_n := \sum_{i=n}^{k_n} \alpha_i^n \psi_i$ we have that for all $\omega \in \Omega_1$ and all $y \in \mathcal{Y}_1$

$$\lim_{n\to\infty} \tilde{\varphi}_n(\omega) =: \varphi(\omega) \text{ and } \lim_{n\to\infty} \tilde{\psi}_n(y) =: \psi(y) \tag{6.18}$$

exist. Extend these functions to Ω and $\mathcal{Y}$, resp., through

$$\limsup_{n\to\infty} \tilde{\varphi}_n(\omega) =: \varphi(\omega) \text{ and } \limsup_{n\to\infty} \tilde{\psi}_n(y) =: \psi(y). \tag{6.19}$$

Due to the boundedness of $\tilde{\varphi}_n$ the same equalities hold with φ^M and $\tilde{\varphi}_n^M$ in place of φ and $\tilde{\varphi}_n$. Given $m \leq n$ we have for $(f, s) \in r(\llbracket 0, \tau \rrbracket), y \in \mathcal{Y}$

$$\mathbb{1}_{A_m}((f, s), y) \leq \tilde{\varphi}_n^M(f, s) + \tilde{\psi}_n(y),$$

hence $\mathbb{1}_{A_m}((f, s), y) \leq \varphi^M(f, s) + \psi(y)$ and thus also

$$\mathbb{1}_A((f, s), y) \leq \varphi^M(f, s) + \psi(y).$$

Given $\varepsilon > 0$, we can find lower semi-continuous functions $\varphi^\varepsilon \geq \varphi$ and $\psi^\varepsilon \geq \psi$ such that $\mathbb{W}(\varphi^\varepsilon) - \varepsilon/2 < \mathbb{W}(\varphi) = \lim \mathbb{W}(\tilde{\varphi}_n)$ and $\sigma(\psi^\varepsilon) - \varepsilon/2 < \sigma(\psi) = \lim \sigma(\tilde{\psi}_n)$. Therefore we can conclude

$$D(A) \leq \limsup_n D(A_n) + 1/n + \varepsilon.$$

To show continuity from above for compact sets, take a sequence $K_1 \supseteq K_2 \supseteq \ldots$ of compact sets in $S \times \mathcal{Y}$ and put $K = \bigcap_n K_n$. Fix $\varepsilon > 0$. Then there is $(\varphi, \psi) \in \mathsf{DC}''$ such that

$$\int \psi \, d\sigma + \int \varphi \, d\mathbb{W} \leq D(K) + \varepsilon.$$

As $(\varphi, \psi) \in \mathsf{DC}''$ it holds that $K \subseteq \{\varphi^M + \psi \geq 1\}$. At the additional cost of ε we can find two lower semi-continuous functions $\varphi^\varepsilon := (\varphi + \varepsilon) \wedge 1 \geq \varphi$ and $\psi^\varepsilon := (\psi + \varepsilon) \wedge 1 \geq \psi$ such that $\mathbb{W}(\varphi^\varepsilon) + \sigma(\psi^\varepsilon) \leq \mathbb{W}(\varphi) + \sigma(\psi) + 2\varepsilon$ and $K \subseteq \{(\varphi^\varepsilon)^M + \psi^\varepsilon > 1\}$. By lower semi-continuity, $\{(\varphi^\varepsilon)^M + \psi^\varepsilon > 1\}$ is open. Hence, there is an N such that for all $n \geq N$ we must have $\mathbb{1}_{K_n} \leq (\varphi^\varepsilon)^M + \psi^\varepsilon$. This implies that

$$D(K_n) \leq D(K) + 2\varepsilon,$$

proving the claim. $\qquad\square$

Recall that for a stopping time ρ we denote by $\bar{\rho}(d\omega, dt)$ the measure on $C(\mathbb{R}_+) \times \mathbb{R}_+$ given by $\delta_{\rho(\omega)}(dt)\mathbb{W}(d\omega)$.

Lemma 6.14. *Let $K \subseteq r(\llbracket 0, \tau \rrbracket) \times \mathcal{Y}$ and assume that $\sup_{\pi \in \mathsf{JOIN}(\tau, \sigma)}(r \otimes \mathrm{Id})(\pi)(K) < 1/2$. Then*

$$1/2 \sup_{\pi \in \mathsf{JOIN}(\tau, \sigma)} (r \otimes \mathrm{Id})(\pi)(K) \leq \inf_{(D,A) \in \mathsf{Cov}(K)} \left(r(\bar{\tau})(D) + \sigma(A) \right) \leq 2 \sup_{\pi \in \mathsf{JOIN}(\tau, \sigma)} (r \otimes \mathrm{Id})(\pi)(K),$$

where

$$\mathsf{Cov}(K) = \{D \subseteq S \text{ open and right complete}, A \subseteq \mathcal{Y} : K \subseteq (D \times \mathcal{Y}) \cup (S \times A)\}.$$

---

[7]We emphasize that a lower-semi-continuous function φ gives rise to an S-lower semi-continuous martingale.

Proof. We will apply the previous lemma. Clearly, $P^{\leq 1}(\mathbb{1}_K) = \sup_{\pi \in \mathsf{JOIN}(\tau,\sigma)}(r \otimes \mathrm{Id})(\pi)(K)$. We have to show that

$$D(K) \approx \inf_{(D,A) \in \mathsf{Cov}(K)} \left(r(\bar{\tau})(D) + \sigma(A) \right).$$

To this end take $(\varphi, \psi) \in \mathsf{DC}''$. As the cost function is $\{0,1\}$-valued, the dual constraint

$$\mathbb{1}_K((f,s),y) \leq \varphi^M(f,s) + \psi(y)$$

implies that

$$K \subseteq (\{(f,s) : \varphi^M(f,s) \geq 1/2\} \times \mathcal{Y}) \cup (S \times \{y : \psi(y) \geq 1/2\}).$$

Hence, on the cost of a factor 2 we can replace ψ by the indicator of a set $A \subseteq \mathcal{Y}$: first, given ψ we can safely replace it by $\tilde{\psi} = \psi \wedge 1$ because $\tilde{\psi}$ has smaller expectation and at least as good covering properties as ψ. Then just take $A = \{\tilde{\psi} \geq 1/2\}$. Obviously, $1/2\sigma(A) \leq \sigma(\tilde{\psi}) \leq \sigma(A)$.

Let us turn our attention to the set $E = \{(f,s) : \varphi^M(f,s) \geq 1/2\}$. Using our assumption on π, we may assume that $\varphi^M(0,0) < 1/2$, and choose $1/2 - \varphi^M(0,0) > \varepsilon > 0$. Define the right complete and open set

$$D = \bigcup_s \left\{ (g,t) : t > s, \varphi^M(g_{\upharpoonright[0,s]}) > 1/2 - \varepsilon \right\}.$$

Due to lower semi-continuity we have $E \subseteq D$. Therefore, $(D,A) \in \mathsf{Cov}(K)$.

By Lemma 4.8, D defines an $\mathcal{F}^+$-stopping time κ such that $r^{-1}(D) =]\!]\kappa, \infty[\![$. Hence,

$$r(\bar{\tau})(D) = \bar{\tau}(]\!]\kappa, \infty[\![) = \mathbb{W}(\kappa < \tau).$$

By Proposition 4.12, there is a continuous martingale ζ which almost surely equals φ^M at all stopping times. This allows us to deduce that

$$\mathbb{E}[\varphi_0^M] = \mathbb{E}[\varphi_\kappa^M] = \mathbb{E}[\zeta_\kappa] \geq (1/2 - \varepsilon)\,\mathbb{W}(\kappa < \tau).$$

As $\varepsilon > 0$ is arbitrary, this implies

$$\inf_{(D,A) \in \mathsf{Cov}(K)} \left(r(\bar{\tau})(D) + \sigma(A) \right) \leq 2 \sup_{\pi \in \mathsf{JOIN}(\tau,\sigma)} (r \otimes \mathrm{Id})(\pi)(K).$$

To show the other inequality, fix $\varepsilon > 0$ and identify D with an $\mathcal{F}^+$-stopping time κ as above and take an open set $O \supseteq \{\kappa < \tau\}$ satisfying $\mathbb{W}(O) \leq 2\mathbb{W}(\kappa < \tau) + \varepsilon = 2r(\bar{\tau})(D) + \varepsilon$. Define a martingale $\bar{\varphi}^M$ by putting

$$\bar{\varphi}(\omega) = \mathbb{1}_O(\omega).$$

As O is open, $\bar{\varphi}$ is lower semi-continuous (and $\bar{\varphi}^M$ is an S- lower semi-continuous martingale) and we have $\mathbb{E}\bar{\varphi} \leq 2r(\bar{\tau})(D) + \varepsilon$. Take an open set $\hat{A} \supseteq A$ satisfying $\sigma(\hat{A}) \leq 2\sigma(A) + \varepsilon$. Then, $(\bar{\varphi}, \mathbb{1}_{\hat{A}}) \in \mathsf{DC}''$ and

$$1/2 \sup_{\pi \in \mathsf{JOIN}(\tau,\sigma)} (r \otimes \mathrm{Id})(\pi)(K) \leq \inf_{(D,A) \in \mathsf{Cov}(K)} \left(r(\bar{\tau})(D) + \sigma(A) \right) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we can conclude. $\qquad\square$

6.2.3. *On Proposition 6.7 — Conclusion of the proof.* Recall that we assume that the stopping time τ is smaller than or equal to some number t_0.

Proof of Proposition 6.7. By Lemma 6.14, for each $\varepsilon > 0$ there exist a right complete set $D \subseteq S$ and a set $N \subseteq \mathcal{Y}$ such that $B \subseteq (D \times \mathcal{Y}) \cup (S \times N)$ and $r(\bar{\tau})(D) + \sigma(N) \leq 2\varepsilon$.

Fix $\eta > 0$ and pick for each k some right complete set $D_k \subseteq S$ and a set $N_k \subseteq \mathcal{Y}$ such that $B \subseteq (D_k \times \mathcal{Y}) \cup (S \times N_k)$ and $r(\bar{\tau})(D_k) + \sigma(N_k) \leq \eta 2^{-k}$. Then setting $\tilde{D} = \bigcap_k D_k$, which is still right complete, we get

$$B \subseteq \bigcap_k \left((D_k \times \mathcal{Y}) \cup \left(S \times \left(\bigcup_j N_j \right) \right) \right) \subseteq \left(\tilde{D} \times \mathcal{Y} \right) \cup \left(S \times \left(\bigcup_j N_j \right) \right).$$

This shows that D, N can be chosen so that $r(\bar{\tau})(D) = 0$ and $\sigma(N) < \varepsilon$, for any $\varepsilon > 0$. Similarly, taking a sequence of such right complete sets D_k' and sets N_k' such that $r(\bar{\tau})(D_k') = 0$ and $\sigma(N_k') < \eta 2^{-k}$, we see that

$$B \subseteq \bigcup_k \left(D_k' \times \mathcal{Y} \right) \cup \left(S \times \left(\bigcap_l N_l' \right) \right).$$

The desired conclusion follows upon setting $D = \bigcup_k D'_k$ and $N = \bigcap_k N'_k$. $\qquad\square$

6.3. **A secondary minimization result.** In certain cases, in order to resolve possible non-uniqueness of a minimizer, it will be useful to identify particular solutions as the solution not only to a primary optimization result, but also as the unique optimizer within this class of a second minimization problem. To this end, we begin by making the following definition: suppose $\gamma : S \to \mathbb{R}$ is a Borel-measurable function, we write $\mathsf{Opt}(\gamma, \mu)$ for the set of optimizers of $P_\gamma(\mu)$. Observe that, when $P_\gamma(\mu)$ is finite and the map $\pi \mapsto \int \gamma \, d\pi$ is lower semi-continuous the set $\mathsf{Opt}(\gamma, \mu)$ is a closed subset of $\mathsf{RST}(\mu)$, and hence also compact.

Theorem 6.15. *Let $\gamma, \tilde{\gamma}$ be Borel measurable functions on S. Suppose that $\mathsf{Opt}(\gamma, \mu) \neq \emptyset$ and that $\nu \in \mathsf{Opt}(\gamma, \mu)$ is an optimizer of*

$$P_{\tilde{\gamma}|\gamma}(\mu) = \inf_{\tilde{\nu} \in \mathsf{Opt}(\gamma,\mu)} \int \tilde{\gamma} \, d\tilde{\nu}. \tag{6.20}$$

Then there exists a Borel set $\Gamma \subseteq S$ such that $r(\nu)(\Gamma) = 1$ and

$$\mathsf{SG_2}^\nu \cap (\Gamma^< \times \Gamma) = \emptyset, \tag{6.21}$$

where

$$\mathsf{SG_2}^\nu = \mathsf{SG}^\nu \cup \big\{ ((f,s),(g,t)) : f(s) = g(t), \tag{6.22}$$

$$\int \gamma(f \oplus \omega_{\upharpoonright[0,r]}, s+r) \, d\nu^{(f,s)}(\omega,r) + \gamma(g,t) = \gamma(f,s) + \int \gamma(g \oplus \omega_{\upharpoonright[0,r]}, t+r) \, d\nu^{(f,s)}(\omega,r),$$

$$\int \tilde{\gamma}(f \oplus \omega_{\upharpoonright[0,r]}, s+r) \, d\nu^{(f,s)}(\omega,r) + \tilde{\gamma}(g,t) > \tilde{\gamma}(f,s) + \int \tilde{\gamma}(g \oplus \omega_{\upharpoonright[0,r]}, t+r) \, d\nu^{(f,s)}(\omega,r) \big\}.$$

We will show that $(r \otimes r)(\pi)(\mathsf{SG_2}^\nu) = 0$ for all $\pi \in \mathsf{JOIN}(\tau_n, \nu)$ where τ_n is defined as in the proof of Theorem 6.5. Then the very same proof as for Theorem 6.5 applies also in the present situation. Hence, the result follows immediately from the following straighforward variant of Proposition 6.6.

Proposition 6.16. *Let ν be a randomized stopping time which minimizes (6.20). Assume that $\pi \in \mathsf{JOIN}(\tau, \nu)$ (where τ can be arbitrary) satisfies*

$$H^{\mathsf{proj}_X(\pi)}(f,s) > 0 \quad \Longrightarrow \quad H^\nu(f,s) > 0 \quad \text{for } (f,s) \in S.$$

Then we have $(r \otimes r)(\pi)(\mathsf{SG_2}^\nu) = 0$.

Proof. As $\nu \in \mathsf{Opt}(\gamma, \mu)$ we only have to show that $(r \otimes r)(\pi)(\mathsf{SG_2}^\nu \setminus \mathsf{SG}^\nu) = 0$ by Proposition 6.6. However, by the very same construction as in the proof of Proposition 6.6 we can again argue by contradiction proving the claim. Indeed, we only have to evaluate the integrals of γ and $\tilde{\gamma}$ as in (6.13) and (6.14), sum them up, use the assumptions and derive a contradiction. $\qquad\square$

7. Embeddings in abundance

In this section, we will show that all existing solutions to (OptSEP) can be established by Theorem 6.5. Moreover, we will give further examples to demonstrate how new embeddings as well as higher dimensional versions of classical embeddings can be constructed using the monotonicity principle.

7.1. **A probabilistic interpretation.** In this section we briefly recall some of the key results which have been obtained so far and provide an interpretation in probabilistic terms. To move closer to classical probabilistic notions, we modify slightly our previous notation and consider again a Brownian motion B on some generic probability space. For a function $\gamma : S \to \mathbb{R}$ which is Borel, $\gamma((B_s)_{s \leq t}, t) = \gamma_t$ corresponds to a stochastic process, optional w.r.t. to the natural filtration generated by B (cf. Proposition 4.4). The value of our optimization problem

$$P_\gamma = \inf\{\mathbb{E}[\gamma_\tau] : \tau \text{ solves (SEP)}\} \tag{7.1}$$

does not depend on the particular choice of the underlying probability space, provided that $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ is sufficiently rich that it supports a Brownian motion $(B_t)_{t \geq 0}$ starting at 0, and an independent, uniformly distributed random variable Y, which is $\mathcal{F}_0$-measurable. We write $\mathcal{F}^n$ for the natural filtration of B. We will from now on assume that we are working in this setting.

Our first result (Theorem 1.1 / Proposition 5.1) says that, for a process $(\gamma_t)_{t \geq 0}$ which is suitably continuous and bounded from below, the infimum in (7.1) is attained by a stopping time $\hat{\tau}$. Moreover, our duality result (Theorem 5.2) says that we have:

$$\inf \{\mathbb{E}[\gamma_\tau] : \tau \text{ solves (SEP)}\} = \sup_{\psi, (M_s)_{s \geq 0}} \int \psi(x) \, d\mu(x) + M_0$$

where the supremum is taken over continuous functions ψ and $\mathcal{F}^n$-martingales M_t (satisfying a certain integrability condition) such that $\gamma_t \geq \psi(B_t) + M_t$.

Our main contribution is the monotonicity principle which describes a given optimizer $\hat{\tau}$ of P_γ in "geometric terms". The version we state here is weaker than what we have actually proved above (cf. Theorem 6.5) but easier to formulate and still sufficient for our intended applications.

To define the set of stop-go pairs, we need one further notion. (See also Figure 4). Set

$$\gamma^{\oplus(f,s)}(h) := \gamma(f \oplus h). \tag{7.2}$$

Then (f, g) constitutes a *stop-go pair*, written $(f, g) \in \mathsf{SG}$, iff $f(s) = g(t)$ and for *every* stopping time σ satisfying $0 < \mathbb{E}[\sigma] < \infty$

$$\mathbb{E}[(\gamma^{\oplus(f,s)})_\sigma] + \gamma(g,t) \quad > \quad \gamma(f,s) + \mathbb{E}[(\gamma^{\oplus(g,t)})_\sigma]. \tag{7.3}$$
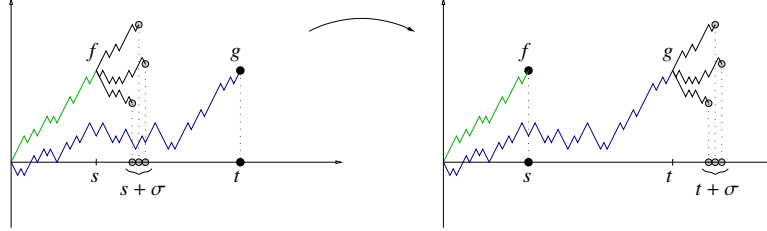

FIGURE 4. The left hand side of (7.3) corresponds to averaging the function γ over the stopped paths on the left picture; the right hand side to averaging the function γ over the stopped paths on the right picture.

We then find a support of $\hat{\tau}$, i.e. a set $\Gamma \subseteq S$ with $\mathbb{P}[((B_t)_{t \leq \hat{\tau}}, \hat{\tau}) \in \Gamma] = 1$, such that

$$\mathsf{SG} \cap (\Gamma^< \times \Gamma) = \emptyset. \tag{7.4}$$

Denote the set of all minimizers of P_γ by Opt_γ and take another Borel function $\tilde{\gamma} : S \to \mathbb{R}$. Assume that $\hat{\tau}$ is also a minimizer of the secondary optimization problem

$$P_{\tilde{\gamma}|\gamma} = \inf\{\mathbb{E}[\tilde{\gamma}_\tau] : \tau \in \mathsf{Opt}_\gamma\}.$$

The set of secondary stop-go pairs SG_2 consists of all $(f, g) \in S \times S$ such that for every stopping time σ with $0 < \mathbb{E}[\sigma] < \infty$ we have

$$\mathbb{E}[(\gamma^{\oplus(f,s)})_\sigma] + \gamma(g,t) \quad \geq \quad \gamma(f,s) + \mathbb{E}[(\gamma^{\oplus(g,t)})_\sigma]$$

and the equality

$$\mathbb{E}[(\gamma^{\oplus(f,s)})_\sigma] + \gamma(g,t) \quad = \quad \gamma(f,s) + \mathbb{E}[(\gamma^{\oplus(g,t)})_\sigma] \tag{7.5}$$

implies the inequality

$$\mathbb{E}[(\tilde{\gamma}^{\oplus(f,s)})_\sigma] + \tilde{\gamma}(g,t) \quad > \quad \tilde{\gamma}(f,s) + \mathbb{E}[(\tilde{\gamma}^{\oplus(g,t)})_\sigma]. \tag{7.6}$$

Then, we also may assume that

$$\mathsf{SG}_2 \cap \left(\Gamma^< \times \Gamma \right) = \emptyset. \tag{7.7}$$

Note that $\mathsf{SG}_2 \supseteq \mathsf{SG}$. We say that Γ is γ-monotone if (7.7) is satisfied. We then obtain:

Theorem 7.1 (Monotonicity Principle II)**.** *Let $\gamma : S \to \mathbb{R}$ be Borel measurable, B be a Brownian motion on some stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and $\hat{\tau}$ an optimizer of the primary as well as secondary optimization problem. Then there exists a γ-monotone Borel set $\Gamma \subseteq S$ such that $\mathbb{P}$-a.s.*

$$((B_t)_{t\leq\hat{\tau}}, \hat{\tau}) \in \Gamma . \tag{7.8}$$

(A) The Azéma-Yor construction.   (B) The Jacka construction   (C) The Vallois construction
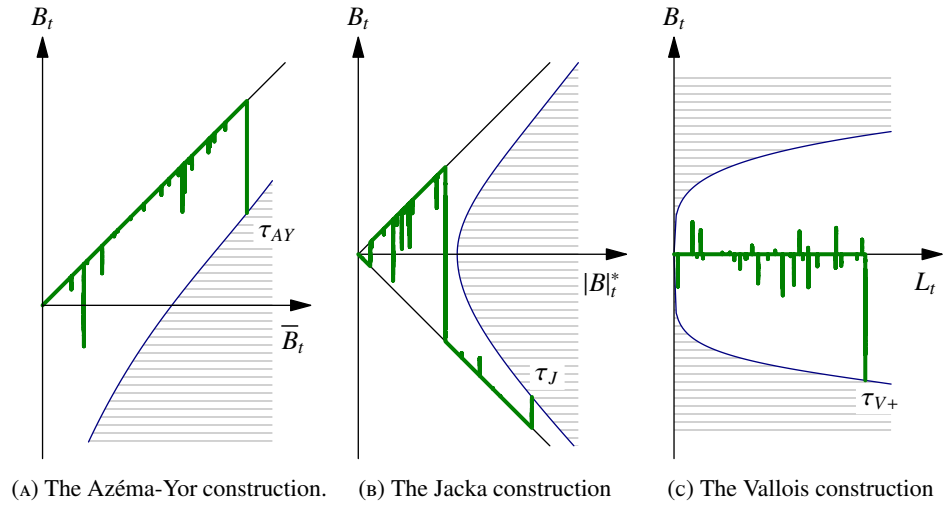
FIGURE 5. Representations of the Azéma-Yor, Vallois and Jacka constructions.

7.2. **Recovering classical embeddings.** In this section we derive a number of classical embeddings as well as establish new embeddings. Figure 5 shows graphical representations of some of these constructions. We highlight the common features of all these pictures: when plotted in an appropriate phase space, the stopping time is the hitting time of a barrier-type set. Identifying the appropriate phase space, and determining the exact structure of the barrier will be the key step in deriving the solutions to (SEP) in this section.

Theorem 7.2 (The Azéma-Yor embedding, cf. [4])**.** *There exists a stopping time τ_{AY} which maximizes*

$$\mathbb{E}\Big[\sup_{t\leq\tau} B_t \Big]$$

over all solutions to (SEP) *and which is of the form $\tau_{AY} = \inf \{t > 0 : B_t \leq \psi(\sup_{s\leq t} B_s)\}$ a.s., for some increasing function ψ.*

For subsequent use, it will be helpful to write, for $(f, s) \in S$, $\bar{f} = \sup_{r\leq s} f(r)$, $\underline{f} = \inf_{r\leq s} f(r)$ and $|f|^* = \sup_{r\leq s} |f(r)|$.

Proof. Fix a bounded and strictly increasing continuous function $\varphi : \mathbb{R}_+ \to \mathbb{R}$ and consider the S-continuous functions $\gamma((f, s)) = -\bar{f}$ and $\tilde{\gamma}((f, s)) = \varphi(\bar{f})(f(s))^2$. By the assumptions on φ and the second moment condition on μ, there exists a minimizer τ_{AY} of P_γ and $P_{\tilde{\gamma}|\gamma}$. Pick, by Theorem 7.1, a γ-monotone set $\Gamma \subseteq S$ supporting τ_{AY}. We claim that

$$\mathsf{SG}_2 \supseteq \{((f, s), (g, t)) \in S \times S : g(t) = f(s), \bar{g} < \bar{f}\}. \tag{7.9}$$

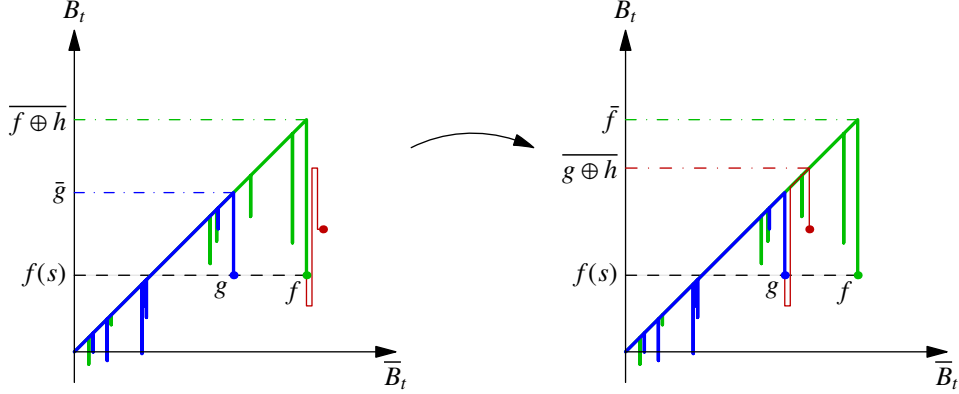This is represented graphically in Figure 6.

FIGURE 6. The stop-go pairs for the Azéma-Yor embedding: on the left, the blue path (g, t) is stopped, and the green path, (f, s), is allowed to continue; a possible continuation, h, being shown in red. On the right hand side we see the effect of allowing g to go and stopping f. In particular, the maximum of g is increased, but the maximum of f stays the same.

Indeed, pick $((f, s), (g, t)) \in S \times S$ with $f(s) = g(t)$ and $\bar{g} < \bar{f}$ and a stopping time σ with positive and finite expectation. Then (7.3) amounts to

$$\mathbb{E}\Big[\bar{f} \vee (f(s) + \bar{B}_\sigma)\Big] + \bar{g} \leq \bar{f} + \mathbb{E}\Big[\bar{g} \vee (g(t) + \bar{B}_\sigma)\Big]$$

with a strict inequality unless $\bar{g} \geq g(t) + \bar{B}_\sigma$ a.s. However in that case (7.5) is trivially satisfied and (7.6) amounts to

$$\mathbb{E}\Big[\varphi(\bar{f})(f(s) + B_\sigma)^2\Big] + \varphi(\bar{g})g(t)^2 > \varphi(\bar{f})f(s)^2 + \mathbb{E}\Big[\varphi(\bar{g})(g(t) + B_\sigma)^2\Big]$$

which holds since $g(t) = f(s)$. Summing up, $((f, s), (g, t)) \in \mathsf{SG} \subseteq \mathsf{SG}_2$ in the former case and $((f, s), (g, t)) \in \mathsf{SG}_2$ in the latter case, proving (7.9).

In complete analogy with the Root embedding presented above we define

$$\mathcal{R}_{\mathrm{CL}} = \{(m, x) : \exists (g, t) \in \Gamma, \bar{g} \leq m, g(t) = x\},$$
$$\mathcal{R}_{\mathrm{OP}} = \{(m, x) : \exists (g, t) \in \Gamma, \bar{g} < m, g(t) = x\},$$

and write $\tau_{\mathrm{CL}}, \tau_{\mathrm{OP}}$ for the first times the process $(\bar{B}_t(\omega), B_t(\omega))$ hits the sets $\mathcal{R}_{\mathrm{CL}}$ and $\mathcal{R}_{\mathrm{OP}}$ respectively. Then we claim $\tau_{\mathrm{CL}} \leq \tau_{AY} \leq \tau_{\mathrm{OP}}$ a.s. Note that $\tau_{\mathrm{CL}} \leq \tau_{AY}$ holds by definition of τ_{CL}. To show the other inequality pick ω satisfying $((B_s(\omega))_{s \leq \tau_{AY}(\omega)}, \tau_{AY}(\omega)) \in \Gamma$ and assume for contradiction that $\tau_{\mathrm{OP}}(\omega) < \tau_{AY}(\omega)$. Then there exists $s \in [\tau_{\mathrm{OP}}(\omega), \tau_{AY}(\omega))$ such that $f := (B_r(\omega))_{r \leq s}$ satisfies $(\bar{f}, f(s)) \in \mathcal{R}_{\mathrm{OP}}$. Since $s < \tau_{AY}(\omega)$ we have $f \in \Gamma^<$. By definition of $\mathcal{R}_{\mathrm{OP}}$, there exists $(g, t) \in \Gamma$ such that $f(s) = g(t)$ and $\bar{g} < \bar{f}$, yielding a contradiction.

Finally, we define

$$\psi_0(m) = \sup\{x : \exists (m, x) \in \mathcal{R}_{\mathrm{CL}}\}.$$

It follows from the definition of $\mathcal{R}_{\mathrm{CL}}$ that $\psi_0(m)$ is increasing, and we define the right-continuous function $\psi_+(m) = \psi_0(m+)$, and the left-continuous function $\psi_-(m) = \psi_0(m-)$. It follows from the definitions of τ_{OP} and τ_{CL} that:

$$\tau_+ := \inf\{t \geq 0 : B_t \leq \psi_+(\bar{B}_t)\} \leq \tau_{\mathrm{CL}} \leq \tau_{\mathrm{OP}} \leq \inf\{t \geq 0 : B_t < \psi_-(\bar{B}_t)\} =: \tau_-.$$

It is then easily checked that $\tau_- = \tau_+$ a.s., and the result follows on taking $\psi = \psi_+$. □

Theorem 7.3 (The Jacka Embedding, cf. [28]). *Let $\varphi : \mathbb{R}_+ \to \mathbb{R}$ be a bounded, strictly increasing right-continuous function. There exists a stopping time τ_J which* maximizes

$$\mathbb{E}\Big[\varphi\Big(\sup_{t \leq \tau} |B_t|\Big)\Big]$$

over all solutions to (SEP)*, and which is of the form*

$$\tau_J = \inf\left\{t > 0 : B_t \geq \gamma_-\left(\sup_{s \leq t}|B_s|\right) \text{ or } B_t \leq \gamma_+\left(\sup_{s \leq t}|B_s|\right)\right\}$$

a.s., for some functions γ_+, γ_-*, where* γ_- *is decreasing, and* $\gamma_+(y) \geq \gamma_-(y)$ *for all* $y > y_0$*,* $\gamma_-(y) = -\gamma_+(y) = \infty$ *for* $y < y_0$*, some* $y_0 \geq 0$*.*

Proof. The proof runs along similar lines to the proof of Theorem 7.2, once we identify

$$\mathsf{SG}_2 \supseteq \{((f, s), (g, t)) \in S \times S : f(s) = g(t), |f|^* > |g|^*\}$$

define

$$\mathcal{R}_{\mathrm{CL}} = \{(m, x) : \exists (g, t) \in \Gamma, |g|^* \leq m, g(t) = x\}$$
$$\mathcal{R}_{\mathrm{OP}} = \{(m, x) : \exists (g, t) \in \Gamma, |g|^* < m, g(t) = x\},$$

and then take

$$\gamma_-(m) = \inf\{x : \exists (m, x) \in \mathcal{R}_{\mathrm{CL}}\}$$
$$\gamma_+(m) = \sup\{x : \exists (m, x) \in \mathcal{R}_{\mathrm{CL}}\}. \qquad \square$$

Remark 7.4. We observe that both the results hold for one-dimensional Brownian motion with an arbitrary starting distribution λ satisfying the usual convex ordering condition.

Theorem 7.5 (The Perkins Embedding, cf. [41]). *Suppose* $\mu(\{0\}) = 0$. *Let* $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ *be a bounded function which is strictly increasing and left-continuous in both arguments. There exists a stopping time* τ_P *which* minimizes

$$\mathbb{E}\left[\varphi\left(\sup_{t \leq \tau}B_t, -\inf_{t \leq \tau}B_t\right)\right]$$

over the set $\mathsf{RST}(\mu)$ *and which is of the form* $\tau_P = \inf\{t > 0 : B_t \notin (\gamma_+(\bar{B}_t), \gamma_-(\underline{B}_t))\}$*, for some decreasing functions* γ_+ *and* γ_- *which are left- and right-continuous respectively.*

Proof. Fix a bounded and strictly increasing continuous function $\tilde{\varphi} : \mathbb{R}_+^2 \to \mathbb{R}$ and consider the S-continuous functions $\gamma((f, s)) = \varphi(\bar{f}, -\underline{f})$ and $\tilde{\gamma}((f, s)) = -(f(s))^2\tilde{\varphi}(\bar{f}, -\underline{f})$. By the assumptions on φ and $\tilde{\varphi}$, there exists a minimizer τ_P of P_γ and $P_{\tilde{\gamma}|\gamma}$. Pick, by Theorem 7.1 a γ-monotone set $\Gamma \subseteq S$ supporting τ_P.

By a similar argument to that given in the proof of Theorem 7.2 we can show

$$\mathsf{SG}_2 \supseteq \{((f, s), (g, t)) \in S \times S : f(s) = g(t), (\bar{f}, -\underline{f}) < (\bar{g}, -\underline{g})\},$$

where $(\bar{f}, -\underline{f}) < (\bar{g}, -\underline{g})$ is to be understood in the partial oder of $\mathbb{R}^2$. Observe from the fact that $\mu(\{0\}) = 0$ that Γ only contains points such that $\underline{g} < 0 < \bar{g}$. Next, we show that Γ contains no points of the type: $\{(g, t) : \underline{g} < g(t) < \bar{g}\}$; since if there were such a point (g, t) with $g(t) = x$ say, then we must also have passed through x on the way to set the most recent extremum (i.e. either between setting the current minimum and the current maximum, or vice versa). Then there exists $(f, s) \in \Gamma^<$ such that $f(s) = x$ and $\varphi(\bar{f}, -\underline{f}) < \varphi(\bar{g}, -\underline{g})$; hence $((f, s), (g, t)) \in \mathsf{SG}_2$.

Now consider the sets:

$$\mathcal{R}_{\mathrm{CL}} = \left\{(m, x) : \exists (g, t) \in \Gamma, g(t) = x = \underline{g}, \bar{g} \geq m\right\} \cup \left\{(x, i) : \exists (g, t) \in \Gamma, g(t) = x = \bar{g}, \underline{g} \leq i\right\}$$
$$= \underline{\mathcal{R}}_{\mathrm{CL}} \cup \bar{\mathcal{R}}_{\mathrm{CL}}$$
$$\mathcal{R}_{\mathrm{OP}} = \left\{(m, x) : \exists (g, t) \in \Gamma, g(t) = x = \underline{g}, \bar{g} > m\right\} \cup \left\{(x, i) : \exists (g, t) \in \Gamma, g(t) = x = \bar{g}, \underline{g} < i\right\}$$
$$= \underline{\mathcal{R}}_{\mathrm{OP}} \cup \bar{\mathcal{R}}_{\mathrm{OP}},$$

and their respective hitting times, $\tau_{\mathrm{CL}}, \tau_{\mathrm{OP}}$. It is immediate that $\tau_{\mathrm{CL}} \leq \tau_P$ a.s., and an essentially identical argument to that used in the proof of Theorem 7.2 gives $\tau_P \leq \tau_{\mathrm{OP}}$ a.s.

We now set

$$\gamma_+(m) = \sup\{x < 0 : (m, x) \in \underline{\mathcal{R}}_{\mathrm{CL}}\}$$
$$\gamma_-(i) = \inf\{x > 0 : (x, i) \in \bar{\mathcal{R}}_{\mathrm{CL}}\}.$$

Then the functions are both clearly decreasing and left- and right-continuous respectively, by definition of the respective sets $\mathcal{R}_{\text{CL}}, \bar{\mathcal{R}}_{\text{CL}}$. Moreover, it is immediate that

$$\tau_{\text{CL}} = \inf\left\{t > 0 : B_t \notin \left(\gamma_+(\bar{B}_t), \gamma_-(\underline{B}_t)\right)\right\},$$

and we deduce that $\tau_{\text{CL}} = \tau_{\text{OP}}$ a.s. by standard properties of Brownian motion. The conclusion follows. $\qquad\square$

Theorem 7.6 (Maximizing the range). *Let $\varphi : \mathbb{R}_+^2 \to \mathbb{R}$ be a bounded function which is strictly increasing and right-continuous in both variables. There exists a stopping time τ_{xr} which* maximizes

$$\mathbb{E}\left[\varphi\left(\sup_{t \le \tau} B_t, -\inf_{t \le \tau} B_t\right)\right]$$

over the set $\text{RST}(\mu)$*, and which is of the form*

$$\tau_{xr} = \inf\left\{t > 0 : B_t \in \left(\gamma_-(-\underline{B}_t, \bar{B}_t), \gamma_+(-\underline{B}_t, \bar{B}_t)\right)\right\} \ a.s.,$$

for some right-continuous functions $\gamma_-(i, m)$ decreasing in both coordinates and $\gamma_+(i, m)$ increasing in both coordinates.

Proof. Our primary objective function will be to minimize $\gamma((f, s)) = -\varphi(\bar{f}, -\underline{f})$; observe that this is a lower-semi-continuous function on S. We again introduce a secondary minimization problem: specifically, we consider the functional $\tilde{\gamma}((f, s)) = (f(s))^2 \tilde{\varphi}(\bar{f}, -\underline{f})$ for some bounded continuous and strictly increasing function $\tilde{\varphi} : \mathbb{R}_+^2 \to \mathbb{R}$. As above, by the assumptions on φ and $\tilde{\varphi}$ there exist a minimizer of P_γ and $P_{\tilde{\gamma}|\gamma}$ denoted by τ_{xr}. Pick, by Theorem 7.1 a γ-monotone set $\Gamma \subseteq S$ supporting τ_{xr}.

By a similar argument to that given in the proof of Theorem 7.2 we can show

$$\text{SG}_2 \supseteq \{((f, s), (g, t)) \in S \times S : f(s) = g(t), (\bar{f}, -\underline{f}) > (\bar{g}, -\underline{g})\},$$

where again $(\bar{f}, -\underline{f}) > (\bar{g}, -\underline{g})$ is to be understood in the partial order of $\mathbb{R}^2$. We put $I(\underline{b}, \bar{b}) := \text{conv}(\{g(t) : (g, t) \in \Gamma, \underline{g} = \underline{b}, \bar{g} = \bar{b}\})$, where conv denotes the convex hull. Moreover, we set $\gamma_-(-\underline{b}, \bar{b}) = \min\{x : x \in I(\underline{b}, \bar{b})\}$ and $\gamma_+(-\underline{b}, \bar{b}) = \max\{x : x \in I(\underline{b}, \bar{b})\}$. If $\gamma_-(-\underline{b}, \bar{b}) = \underline{b}$ we set $\gamma_-(-\underline{d}, \bar{d}) = \underline{d}$ for all $\underline{d} \le \underline{b}$ and $\bar{d} \ge \bar{b}$ and anologously for γ_+.

We claim that γ_+ is increasing in $\bar{b}$ and $-\underline{b}$ and γ_- is decreasing in $\bar{b}$ and in $-\underline{b}$, i.e. $I(\underline{b}, \bar{b}) \subseteq I(\underline{d}, \bar{d})$ if $\underline{d} \le \underline{b}$ and $\bar{b} \le \bar{d}$. Assume the contrary, then with the notation as before there is $x \in I(\underline{b}, \bar{b}) \setminus I(\underline{d}, \bar{d})$. W.l.o.g. we can assume that $x = \gamma_-(-\underline{b}, \bar{b}) > \underline{b}$. Then there is $(f, s) \in \Gamma^<, (g, t) \in \Gamma$ with $\underline{g} = \underline{b}, \bar{g} = \bar{b}, g(t) = \gamma_-(-\underline{b}, \bar{b}) = f(s)$ and $\underline{f} = \underline{d}, \bar{f} = \bar{d}$, in particular $\varphi(\bar{f}, -\underline{f}) > \varphi(\bar{g}, -\underline{g})$ so that $(f, g) \in \text{SG}_2$.

Set

$$\mathcal{R}_{\text{CL}} = \{(\underline{b}, \bar{b}, b) : b \in I(\underline{b}, \bar{b})\}$$

$$\mathcal{R}_{\text{OP}} = \{(\underline{b}, \bar{b}, b) : \exists \underline{d} \ge \underline{b}, \bar{d} \le \bar{b} \text{ one inequality being strict, s.t. } b \in I(\underline{d}, \bar{d})\}$$

with respective hitting times τ_{CL} and τ_{OP}. In a similar manner to the Azéma-Yor embedding we can deduce that $\tau_{\text{CL}} \le \tau_{rx} \le \tau_{\text{OP}}$ a.s. and also that $\tau_{\text{CL}} = \tau_{\text{OP}}$ a.s. The conclusion follows upon noting that

$$\tau_{\text{CL}} = \inf\left\{t > 0 : B_t \in \left(\gamma_-(-\underline{B}_t, \bar{B}_t), \gamma_+(-\underline{B}_t, \bar{B}_t)\right)\right\}. \qquad\square$$

Remark 7.7. Considering the last argument we see that looking for minimizers of the range the picture turns inside out. The stopping region will be of the form $(-\infty, a) \cup (b, \infty)$ and we can directly deduce that we only stop in a minimum or a maximum, i.e. the three dimensional picture reduces to a two dimensional picture and we are back in the Perkins case. This symmetry is similar to the symmetry in the Root and Rost embedding.

Remark 7.8. We observe that, in the case of Theorem 7.6, the characterization provided would not appear to be sufficient to identify the functions γ_+, γ_- given the measure μ. This is in contrast to the constructions of Azéma-Yor, Perkins and Jacka, where knowledge of the form of the embedding is sufficient to identify the corresponding stopping rule. Consider the Azéma-Yor embedding: from Theorem 7.2 it is clear that if we stop with

maximum above s, then we must hit s before stopping, and there exists $y = \psi(s)$ such that we stop at or above y if our maximum is above s. Moreover, we only stop above y if our maximum is at least s. We conclude that, conditional on hitting s before stopping, we must embed μ restricted to $[y, \infty)$ (with some more care needed if there are atoms of μ). However there is a unique y such that this distribution has mean s, and this y must then be $\psi(s)$.

On a more abstract level, uniqueness of barrier type embeddings in a two dimensional phase space can be seen as a consequence of Loynes' argument [34]. More precisely, let A_t be some continuous process and suppose that τ_1 and τ_2 denote the times when (A_t, B_t) hits a closed barrier-type set R_1 resp. R_2. If $\mathbb{E}[\tau_1], \mathbb{E}[\tau_2] < \infty$ and both stopping times embed the same measure, the argument presented in Remark 2.2 shows that $\tau_1 = \tau_2$.

7.2.1. *The Vallois-embedding and optimizing functionals of local time.* In this section we shall determine the stopping rule which solves

$$\inf\{\mathbb{E}[h(L_\tau)] : \tau \text{ solves (SEP)}\}, \tag{7.10}$$

where L denotes the local time of Brownian motion in 0 and h is a convex or concave function.

A large part of the argument is virtually identical to the argument which we used in the previous section. The most involved part will in fact be to show that the problem (7.10) admits a maximizer. As mentioned below Definition 4.3, L is not S- continuous. Nevertheless we will prove the following result.

Lemma 7.9. *Let $\xi_n, n \geq 1, \xi \in \mathsf{RST}(\mu)$ and assume that $\xi_n \to \xi$ weakly. Then $L_{\xi_n} \to L_\xi$ weakly. In fact, if ρ_n, ρ are the representations of ξ_n, ξ on $\bar{\Omega}$, then $\bar{L}_{\rho_n} \to \bar{L}_\rho$ in $L^1(\bar{\Omega}, \mathbb{P} \otimes \mathcal{L})$.*

We first give a simple result on the connection between convergence of stopping times in $\mathsf{RST}(\mu)$ and their representations.

Proposition 7.10. *Let ξ_n, ξ, ρ_n and ρ be as in Lemma 7.9. Then $\xi_n \to \xi$ weakly iff $\rho_n \to \rho$ in probability.*

Proof. Let $X \in C_b(C(\mathbb{R}_+) \times \mathbb{R}_+)$. Recall that

$$\int X_t(\omega) \, d\xi(\omega, t) = \int \mathbb{W}(d\omega) \int \xi_\omega(dt) X_t(\omega) = \int \mathbb{W}(d\omega) \int X(\omega, t) \, dA_t^\xi(\omega) =$$
$$= \int \mathbb{W}(d\omega) \int \mathcal{L}(dx) X_{\rho(x,\omega)}(\omega) = \bar{\mathbb{E}} X_\rho,$$

and hence

$$\int X_t(\omega) \, d(\xi - \xi_n)(\omega, t) = \int \mathbb{W}(d\omega) \int \mathcal{L}(dx) \left[X_{\rho(x,\omega)}(\omega) - X_{\rho_n(x,\omega)}(\omega)\right].$$

Considering processes which depend only on the time t but not ω, i.e. $X_t(\omega) = X_t$, we obtain that $\xi_n \to \xi$ weakly implies that $\rho_n \to \rho$ in probability. Conversely, if $\rho_n \to \rho$ in probability under $\bar{\mathbb{P}}$, then also $\rho_n \to \rho$ almost surely along some subsequence of every subsequence. By dominated convergence, $\xi_n \to \xi$ weakly. □

Proof of Lemma 7.9. As a consequence of Proposition 7.10 we have that $\rho_n \wedge \rho \to \rho, \rho_n \vee \rho \to \rho$. Note also that for every minimal embedding $\xi' \in \mathsf{RST}(\mu')$, $\mathbb{E}_{\xi'} L = \int |x| \, d\mu'(x)$. Write μ_n for the law embedded by $\rho_n \wedge \rho$. Then $\mu_n \to \mu$ weakly, and $\bar{L}_{\rho_n \wedge \rho} \leq \bar{L}_\rho$, so (using Lemma 4.25) $\bar{\mathbb{E}} \bar{L}_{\rho_n \wedge \rho} = \int |x| \, d\mu_n \to \int |x| \, d\mu$ and hence $\bar{\mathbb{E}} \bar{L}_{\rho_n \wedge \rho} \to \bar{\mathbb{E}} \bar{L}_\rho$. This implies that $\bar{L}_{\rho_n \wedge \rho} \to \bar{L}_\rho$ in $L^1(\bar{\Omega}, \mathbb{P} \otimes \mathcal{L})$. Since $\bar{L}_{\rho_n \vee \rho} + \bar{L}_{\rho_n \wedge \rho} = \bar{L}_\rho + \bar{L}_{\rho_n}$ we also find that $\bar{\mathbb{E}} \bar{L}_{\rho_n \vee \rho} = \bar{\mathbb{E}} \left[\bar{L}_{\rho_n} + (\bar{L}_\rho - \bar{L}_{\rho_n})_+\right] = \bar{\mathbb{E}} \bar{L}_\rho + \bar{\mathbb{E}} \left[(\bar{L}_\rho - \bar{L}_{\rho_n})_+\right] \to \bar{\mathbb{E}} \bar{L}_\rho$, where we used that $\xi_n, \xi \in \mathsf{RST}(\mu)$. Thus $\bar{L}_{\rho_n \vee \rho} \to \bar{L}_\rho$ in $L^1(\bar{\Omega}, \mathbb{P} \otimes \mathcal{L})$. Combining these results, we see that $\bar{L}_{\rho_n} \to \bar{L}_\rho$ in $L^1(\bar{\Omega}, \mathbb{P} \otimes \mathcal{L})$. □

Corollary 7.11. *Let $h : [0, \infty) \to \mathbb{R}$ be a bounded, strictly concave or convex function. Then there exists an optimizer for (7.10). Moreover, the set $\mathsf{Opt}(\gamma, \mu)$ is closed.*

Proof. If $\xi_n \in \mathsf{RST}(\mu)$ is a sequence of stopping times such that $V^* = \lim \mathbb{E}[h(L_{\xi_n})] = \inf\{\mathbb{E}[h(L_\tau)] : \tau \in \mathsf{RST}(\mu)\}$, by Lemma 7.9 (possibly passing to a subsequence) $\xi = \lim_n \xi_n$ satisfies $\mathbb{E}[h(L_\xi)] = V^*$. Hence, $\mathsf{Opt}(\gamma, \mu)$ is non-empty and closed. □

With these tools, we are now able to show:

Theorem 7.12. *Let $h : [0, \infty) \to \mathbb{R}$ be a bounded, strictly concave function.*

(1) There exists a stopping time τ_{V-} which maximizes

$$\mathbb{E}[h(L_\tau)]$$

over the set $\mathsf{RST}(\mu)$, and which is of the form

$$\tau_{V-} = \inf\{t > 0 : B_t \notin (\varphi_-(L_t), \varphi_+(L_t))\} \ a.s.,$$

for some decreasing, non-negative function φ_+ and increasing, non-positive function φ_-.

(2) There exists a stopping time τ_{V+} which minimizes

$$\mathbb{E}[h(L_\tau)]$$

over the set $\mathsf{RST}(\mu)$, and which is of the form

$$\tau_{V+} = \inf\{t > 0 : B_t \notin (\varphi_-(L_t), \varphi_+(L_t))\} \wedge Z \ a.s.,$$

for some increasing, non-negative function φ_+, decreasing, non-positive function φ_-, and an $\mathcal{F}_0$-measurable random variable $Z \in \{0, \infty\}$ with $\mathbb{P}(Z = 0) = \mu(\{0\})$.

Proof. We consider the second case, the first case being slightly simpler. We will apply Theorem 7.1 to the optimizations corresponding to $\gamma((\omega, t)) = h(L_t(\omega))$ and $\tilde{\gamma}((\omega, t)) = e^{-L_t(\omega)} B_t^2(\omega)$.

By Corollary 7.11, the set $\mathsf{Opt}(\gamma, \mu)$ is non-empty and closed and we can apply Theorem 7.1. Pick a minimizer τ_{V+} and a γ-monotone set $\Gamma \subseteq S$ supporting τ_{V+}.

From Theorem 4.1 and Proposition 4.4 it follows that there is a function $L : S \to \mathbb{R}_+$ such that $L_t = L((B_s)_{s \leq t}, t)$ $\mathbb{P}$-a.s. By a similar argument to the previous cases we can show that

$$\mathsf{SG}_2 \supseteq \{((f, s), (g, t)) \in S \times S : f(s) = g(t), L(f, s) < L(g, t)\}.$$

Define the sets

$$\mathcal{R}_{\mathrm{OP}} = \{(l, x) : \exists (g, t) \in \Gamma, g(t) = x, L((g, t)) > l\},$$
$$\mathcal{R}_{\mathrm{CL}} = \{(l, x) : \exists (g, t) \in \Gamma, g(t) = x, L((g, t)) \geq l\}.$$

It follows immediately that $\tau_{\mathrm{CL}} \leq \tau_{V+}$. We observe that $(l, 0) \notin \mathcal{R}_{\mathrm{OP}}$ for any $l \geq 0$, and $\tau_{V+} \leq \tau_{\mathrm{OP}}$, and it follows that $\tau_{V+}(\Omega \times \{0\}) = \mu(\{0\})$. We consider τ_{V+} on $\{\tau_{V+} \geq \eta\}$, for $\eta > 0$. Write $\tau_{\mathrm{OP}}^\eta = \inf\{t \geq \eta : (L_t, B_t) \in \mathcal{R}_{\mathrm{OP}}\}$ and $\tau_{\mathrm{CL}}^\eta = \inf\{t \geq \eta : (L_t, B_t) \in \mathcal{R}_{\mathrm{CL}}\}$. Then on this set $\tau_{\mathrm{CL}}^\eta \leq \tau_{V+} \leq \tau_{\mathrm{OP}}^\eta$. Moreover, define $\varphi_+(l) = \inf\{x > 0 : (l, x) \in \mathcal{R}_{\mathrm{OP}}\}$ and $\varphi_-(l) = \sup\{x < 0 : (l, x) \in \mathcal{R}_{\mathrm{OP}}\}$. Observe that, since $\tau_{V+} \leq \tau_{\mathrm{OP}}$, if $\mathbb{P}(\tau_{V+} > \eta) > 0$, then $|\varphi_+(\eta) - \varphi_-(\eta)| > 0$. In addition, $\varphi_+(l)$ is clearly right-continuous and increasing, so it must have at most countably many discontinuities, and similarly for $\varphi_-(l)$. We can write

$$\inf\{t \geq \eta : B_t \notin (\varphi_-(L_t), \varphi_+(L_t))\} \leq \tau_{\mathrm{CL}}^\eta \leq \tau_{\mathrm{OP}}^\eta \leq \inf\{t \geq \eta : B_t \notin [\varphi_-(L_t+), \varphi_+(L_t+)]\}$$

and observe that (by standard properties of Brownian motion) the stopping times on the left and right are almost surely equal (since there are at most countably many discontinuities, and $\varphi_+(l)$ and $\varphi_-(l)$ are bounded away from zero on $[\eta, \infty)$). It follows that $\tau_{V+} = \inf\{t \geq \eta : B_t \notin (\varphi_-(L_t), \varphi_+(L_t))\}$ on $\{\tau_{V+} \geq \eta\}$, and since $\eta > 0$ was arbitrary, we get the desired behaviour. □

Remark 7.13. The arguments above extend from local time at 0 to a general continuous additive functional A. Writing L^x for local time in x, A can be represented in the form $A_t := \int_0^t L_s^x \, dm_A(x)$. Let f be a convex function such that $f'' = m_A$ in the sense of distributions. If $\int f \, d\mu < \infty$, then Proposition 7.10 still holds with A in place of L; the above proof is easily adapted to the more general situation.

In this manner, we deduce the existence of optimal solutions to (SEP) for functionals depending on A. By analogy with Theorem 7.12 this can be used to generate (inverse-/cave-) barrier-type embeddings of various kinds. Other generalizations and variants may be considered in a similar manner. We leave specific examples as an exercise for the reader.

Remark 7.14. In Cox and Obłój [12], embeddings are constructed which maximize certain double-exit probabilities: for example, to maximize the probability that both $\bar{B}_\tau \geq \bar{b}$ and $\underline{B}_\tau \leq \underline{b}$, for given levels $\bar{b}$ and $\underline{b}$. In this case, the embedding is no longer naturally viewed as a barrier-type construction; instead, it is natural to characterize the embedding in terms of where the paths with different crossing behaviour for the barriers finish (for example, the paths which only hit the upper level may end up above a certain value, or between two other values). However, it is possible, again using a suitable secondary maximization problem, to show that there exists an optimizer demonstrating the behaviour characterizing the Cox-Obłój embeddings. (Specifically, if we write $H_b((f,s)) = \inf\{t \leq s : f(t) = b\}$, $\underline{H} = H_{\underline{b}} \wedge H_{\bar{b}}$ and $\bar{H} = H_{\underline{b}} \vee H_{\bar{b}}$ then the secondary maximization problem $\tilde{\gamma}((f,s)) = ((f(s) - \underline{H}((f,s)))^2 \mathbb{1}_{\underline{H} \leq s}/2 - ((f(s) - \bar{H}((f,s)))^2 \mathbb{1}_{\bar{H} \leq s}$ is sufficient to rederive the form of these embeddings.)

7.3. **Root and Rost Embeddings in Higher Dimensions.** In this section we consider the Root and Rost constructions of Sections 2.1 and 2.2 in the case of d-dimensional Brownian motion with general initial distribution, for $d \geq 2$. In $\mathbb{R}^d$, since the Brownian motion is transient, it is no longer straightforward to assert the existence of an embedding. In general, [47] gives necessary and sufficient conditions for the existence of an embedding, and without the additional condition that $\mathbb{E}[\tau] < \infty$. In the Brownian case, Rost's conditions for $d \geq 3$ can be written as follows. There exists a stopping time τ such that $B_0 \sim \lambda$ and $B_\tau \sim \mu$ if and only if for all $y \in \mathbb{R}^d$

$$\int u(x,y)\,\lambda(dx) \leq \int u(x,y)\,\mu(dx), \text{ where } u(x,y) = |x-y|^{2-d}. \tag{7.11}$$

However, it is not clear that such a stopping time will satisfy the condition

$$\mathbb{E}[\tau] = 1/d\left(\int |x|^2\,(\mu - \lambda)(dx)\right). \tag{7.12}$$

As a result, it is not straightforward to give simple criteria for the existence of a solution in $\mathsf{RST}(\mu)$.

In the case $d = 2$ it follows from Falkner's results [19] that the Skorokhod problem admits a solution (i.e. $\mathsf{RST}(\mu) \neq \emptyset$) if (7.11) is satisfied for $u(x,y) = -\ln|x-y|$ and then (7.12) applies.

In either case, assuming that we do have a solution satisfying (7.12), then we are able to state the following:

Theorem 7.15. *Suppose* $\mathsf{RST}(\mu)$ *is non-empty. If h is a strictly convex function and $\hat{\tau} \in \mathsf{RST}(\mu)$ maximizes $\mathbb{E}[h(\tau)]$ over $\tau \in \mathsf{RST}(\mu)$ then there exists a barrier $\mathcal{R}$ such that $\hat{\tau} = \inf\{t > 0 : (B_t, t) \in \mathcal{R}\}$ on $\{\hat{\tau} > 0\}$ a.s.*

The proof of this result is much the same as that of Theorem 2.1, except we no longer show that $\tau_{\text{CL}} = \tau_{\text{OP}}$. In higher dimensions with general initial laws, it is easy to construct examples where there are common atoms of λ and μ, but where the size of the atom in λ is strictly larger than the atom of μ. By the transience of the process, it is clear that the optimal (indeed, only) behaviour is to stop mass starting at such a point immediately with a probability strictly between 0 and 1, however the stopping times τ_{CL} and τ_{OP} will always stop either all the mass, or none of this mass respectively. For this reason, we do not say anything about the behaviour of $\hat{\tau}$ when $\hat{\tau} = 0$. Trivially, the above result tells us that the solution of the optimal embedding problem is given by a barrier if there exists a set D such that $\lambda(D) = 1 = \mu(\complement D)$.

Proof of Theorem 7.15. The first part of the proof proceeds similarly to the proof of Theorem 2.1. In particular, the set of stop-go pairs is given by

$$\mathsf{SG} \supseteq \{((f,s),(g,t)) \in S \times S : f(s) = g(t), s > t\}$$

and we define the sets $\mathcal{R}_{\mathrm{CL}}, \mathcal{R}_{\mathrm{OP}}$ and the stopping times $\tau_{\mathrm{CL}}, \tau_{\mathrm{OP}}$ as above. We then fix $\delta > 0$, and consider the set $\{\hat{\tau} \geq \delta\}$. Given $\eta \geq 0$, we define $B_t^{-\eta} = B_{t+\eta}$, for $t \geq -\eta$ and set

$$\tau_{\mathrm{CL}}^{\eta,\delta} = \inf\{t \geq \delta : (t, B_t^{-\eta}) \in \mathcal{R}_{\mathrm{CL}}\}.$$

Then $\tau_{\mathrm{CL}}^{\eta,\delta} \geq \delta$, and for any $\varepsilon > 0$, we can choose $\eta > 0$ sufficiently small that

$$d_{TV}\left(B_\delta^{-\eta}, B_\delta\right) < \varepsilon$$

and hence from the Strong Markov property of Brownian motion, it follows that

$$d_{TV}\left(B_{\tau_{\mathrm{CL}}^{\eta,\delta}}^{-\eta}, B_{\tau_{\mathrm{CL}}^{0,\delta}}\right) < \varepsilon.$$

In particular, the law of $B_{\tau_{\mathrm{CL}}^{\eta,\delta}}^{-\eta}$ converges weakly to the law of $B_{\tau_{\mathrm{CL}}^{0,\delta}}$ as $\eta \to 0$. Thus

$$\tau_{\mathrm{CL}}^{\eta,\delta} = \inf\{t \geq \eta + \delta : (t - \eta, B_t) \in \mathcal{R}_{\mathrm{CL}}\},$$

so $\tau_{\mathrm{CL}}^{\eta,\delta} \geq \tau_R^{0,\delta}$, and moreover, $\tau_{\mathrm{CL}}^{\eta,\delta} \to \tau_{\mathrm{OP}}^{0,\delta}$ a.s. as $\eta \to 0$. Hence, $B_{\tau_{\mathrm{CL}}^{\eta,\delta}}^{-\eta} \to B_{\tau_{\mathrm{OP}}^{0,\delta}}$ in probability, as $\eta \to 0$, so we have weak convergence of the law of $B_{\tau_{\mathrm{CL}}^{\eta,\delta}}^{-\eta}$ to the law of $B_{\tau_{\mathrm{OP}}^{0,\delta}}$, and hence

$$B_{\tau_{\mathrm{OP}}^{0,\delta}} \sim B_{\tau_{\mathrm{CL}}^{0,\delta}}.$$

We now observe that, by an essentially identical argument to that in the proof of Theorem 2.1, we must have $\tau_{\mathrm{CL}}^{0,\delta} \leq \hat{\tau} \leq \tau_{\mathrm{OP}}^{0,\delta}$ on $\{\hat{\tau} \geq \delta\}$. However, in the argument above, we know that $\tau_{\mathrm{CL}}^{0,\delta} \leq \hat{\tau} \leq \tau_{\mathrm{OP}}^{0,\delta}$, and $\tau_{\mathrm{CL}}^{\eta,\delta} \to_{\mathcal{D}} \tau_{\mathrm{CL}}^{0,\delta}$ and $\tau_{\mathrm{CL}}^{\eta,\delta} \to_{\mathcal{D}} \tau_{\mathrm{OP}}^{0,\delta}$ as $\eta \to 0$ (where $\mathcal{D}$ denotes convergence in distribution). It follows that $\tau_{\mathrm{CL}}^{0,\delta} =_{\mathcal{D}} \tau_{\mathrm{OP}}^{0,\delta}$ and hence $\tau_{\mathrm{CL}}^{0,\delta} = \tau_{\mathrm{OP}}^{0,\delta}$ a.s. In particular, $B_{\tau_{\mathrm{CL}}^{0,\delta}} = B_{\tau_{\mathrm{OP}}^{0,\delta}} = B_{\hat{\tau}}$ on $\{\hat{\tau} \geq \delta\}$. Letting $\delta \to 0$ we observe that $\tau_{\mathrm{OP}}^{0,\delta} \to \tau_{\mathrm{OP}}$, and hence the required result holds on taking $\mathcal{R} = \mathcal{R}_{\mathrm{OP}}$. $\qquad\square$

We now consider the generalization of the Rost embedding. We recall that $(\lambda \wedge \mu)(A) := \inf_{B \subseteq A} (\lambda(B) + \mu(A \setminus B))$ defines a measure.

Theorem 7.16. *Suppose λ, μ are measures in $\mathbb{R}^d$ and $\hat{\tau} \in \mathsf{RST}(\mu)$ maximizes $\mathbb{E}[h(\tau)]$ over all stopping times in $\mathsf{RST}(\mu)$, for a convex function $h : \mathbb{R}_+ \to \mathbb{R}$, with $\mathbb{E}[h(\tau)] < \infty$. Then $\mathbb{P}(\hat{\tau} = 0, B_0 \in A) = (\lambda \wedge \mu)(A)$, for $A \in \mathcal{B}(\mathbb{R})$, and on $\{\hat{\tau} > 0\}$, $\hat{\tau}$ is the first hitting time of an inverse barrier.*

Proof. We follow the proof of Theorem 2.3 to recover the set of stop-go pairs given by

$$\mathsf{SG} \supseteq \{((f, s), (g, t)) \in S \times S : f(s) = g(t), s < t\}$$

and the sets $\mathcal{R}_{\mathrm{OP}}$ and $\mathcal{R}_{\mathrm{CL}}$, and their corresponding hitting times $\tau_{\mathrm{OP}}, \tau_{\mathrm{CL}}$. For $0 \leq \eta \leq \delta$, we define in addition the stopping times

$$\tau_{\mathrm{CL}}^{\eta,\delta} = \inf\{t \geq \delta : (t, B_t^\eta) \in \mathcal{R}_{\mathrm{CL}}\},$$
$$\tau_{\mathrm{OP}}^{\eta,\delta} = \inf\{t \geq \delta : (t, B_t^\eta) \in \mathcal{R}_{\mathrm{OP}}\},$$

where $B_t^\eta = B_{t-\eta}$, for $t \geq \eta$.

It follows from an identical argument to that in the proof of Theorem 2.3 that $\tau_{\mathrm{CL}}^{0,\delta} \leq \hat{\tau} \leq \tau_{\mathrm{OP}}^{0,\delta}$ on $\{\hat{\tau} \geq \delta\}$. However, by similar arguments to those used above, we deduce that $\tau_{\mathrm{OP}}^{0,\delta}$ and $\tau_{\mathrm{CL}}^{0,\delta}$ have the same law on $\{\hat{\tau} \geq \delta\}$, and hence that $\hat{\tau} = \tau_{\mathrm{OP}}^{0,\delta}$ on this set, and then by taking $\delta \to 0$, we get $\hat{\tau} = \tau_{\mathrm{OP}}$ on $\{\hat{\tau} > 0\}$.

To see the final claim, we note that trivially $\mathbb{P}(\hat{\tau} = 0, B_0 \in A) \leq (\lambda \wedge \mu)(A)$. If there is strict inequality, then there exist some paths which start at $x \in A$, and paths which stop at x at strictly positive time, constituting a stop-go pair and therefore violating the monotonicity principle. $\qquad\square$

Remark 7.17. We observe that the arguments of Remark 2.2 can be applied again in this context. However, one needs to be a little more careful, since it is necessary to take the *fine closure* of the barriers with respect to the fine topology for the processes (t, B_t). With this modification in place, the argument of Loynes can be easily adapted to show that the

(finely closed versions) of the barriers in Theorems 7.15 and 7.16 are unique in the sense of Remark 2.2.

8. Embedding Feller processes

In this section we discuss which changes are needed to establish the duality result, Theorem 5.2, as well as the monotonicity principle, Theorem 6.5, for continuous Feller processes. In fact, most of our arguments are abstract and do not use any specific structure of the Wiener measure. The relation between the spaces S and $X = C(\mathbb{R}_+) \times \mathbb{R}_+$ as well as the approximation of stopping times rely on abstract theory of stochastic processes and topological properties of S and X. The proof of Theorem 6.5 uses duality theory of optimal transport and Choquet's theorem. Proposition 4.11 together with Lemma 4.8 is very valuable to identify certain hitting times as stopping times. To prove the duality statement we use again duality theory of optimal transport and — crucially — the compactness of $\mathsf{RST}(\mu)/\mathsf{PRST}(\mu)$.

This last point, the compactness of $\mathsf{RST}(\mu)$ and the characterization of minimal stopping times in terms of the expectation $\mathbb{E}_\xi[T] = V < \infty$ is in fact the only point where we use specific properties of Brownian motion (apart from Section 7).

So we assume now that we are given a continuous Feller process $Z = (Z_t)_{t \geq 0}$. As usual we assume Z to be the canonical process on the space of continuous functions. We write $(\mathbb{P}_x)_{x \in \mathbb{R}}$ for the law of the Feller process started in x and $\mathbb{P}$ for the law of the process started with law λ.

We define the set RST (PRST, resp.) as before with $\mathbb{P}$ replacing $\mathbb{W}$. Let $\mu \in \mathcal{P}(\mathbb{R})$. We say that $\xi \in \mathsf{RST}$ is a minimal embedding of μ if the corresponding stopping time ρ (cf. (4.4)) on the enlarged probability space $(\bar{\Omega}, \bar{\mathbb{P}})$ constitutes a minimal embedding, i.e. ρ embeds μ in Z and for any ρ' with $\rho' \leq \rho$ and also embedding μ it holds that $\rho' = \rho$.

Definition 8.1. *For $\mu \in \mathcal{P}(\mathbb{R})$ we define $\mathsf{RST}(\mu)$ to be the set of all minimal randomized stopping times embedding the measure μ.*

As above we will also consider the set $\mathsf{PRST}(\mu)$.

Assumption 8.2. *From now on we* assume *that $\mathsf{RST}(\mu)$ is non-empty, compact, and either:*

(1) That there exists an increasing, $\mathcal{F}^0$-optional process $\zeta : X \to \mathbb{R}$ with $\zeta_s \to \infty$ $\mathbb{P}$-a.s. as $s \to \infty$ such that the following hold true:
  - *For a finite $\xi \in \mathsf{RST}$ with $Z_\xi \sim \mu$ we have $\mathbb{E}_\xi[\zeta_t] < \infty$ if and only if ξ is minimal.*
  - *There is a corresponding S-continuous martingale $X_t = h(Z_t) - \zeta_t$ such that $X_{t \wedge \xi}$ is uniformly integrable for all $\xi \in \mathsf{RST}(\mu)$.*

  or
(2) That $\xi \in \mathsf{RST}$ and $B_\xi \sim \mu$ implies ξ is minimal (i.e. all embeddings are minimal).

Below we will verify that this assumption is satisfied in a number of natural examples. Note that compactness of $\mathsf{RST}(\mu)$ is equivalent to the existence of an increasing and diverging function $G : \mathbb{R}_+ \to \mathbb{R}$ such that

$$\sup_{\xi \in \mathsf{RST}(\mu)} \mathbb{E}_\xi[G(T)] =: V_G < \infty.$$

We first show that (1) of Assumption 8.2 is also relevant in the usual Brownian setup, where it allows us to dispose of the second moment condition.

Proposition 8.3. *Let Z be Brownian motion and assume that λ and μ have first moments and are in convex order. Then Assumption 8.2 (1) holds.*

Proof. By the de la Vallée-Poussin theorem (see e.g. [15, Thm. II 22]) there exists a positive, smooth and symmetric function $F : \mathbb{R} \to \mathbb{R}_+$ with strictly positive, bounded second derivative and $\lim_{x \to \infty} F(x)/x = \infty$ such that $V := \int F(x) \mu(dx) < \infty$. We set

$$\zeta_t(\omega) = 1/2 \int_0^t F''(\omega_s) \, ds$$

and note that by Ito's formula and our conditions on F,

$$X_t = F(Z_t) - 1/2 \int_0^t F''(Z_s) \, ds = F(Z_t) - \zeta_t$$

is an S-continuous martingale.

Note also that in the present Brownian case, it is known that the minimality of a finite stopping time ξ is equivalent to $(Z_{t \wedge \xi})_{t \geq 0}$ being a uniformly integrable martingale. This follows (in the case of a general starting law) from Lemma 12 and Theorem 17 of [10].

Assume now that $Z_\xi \sim \mu$ and that $(Z_{t \wedge \xi})_{t \geq 0}$ is uniformly integrable. Then for each t, $Z_{t \wedge \xi}$ is bounded by μ in the convex order and in particular $\mathbb{E}_{\xi \wedge t}[F(Z_s)] \leq V$. We obtain

$$\mathbb{E}_\xi \zeta_s = \lim_{t \to \infty} \mathbb{E}_{t \wedge \xi}[\zeta_s] = \lim_{t \to \infty} \mathbb{E}_{t \wedge \xi}[F(Z_s)] - \mathbb{E}[F(Z_0)] \leq V - \mathbb{E}[F(Z_0)] < \infty.$$

Next assume that $\mathbb{E}_\xi[\zeta_s] < \infty$. Then $\sup_{t \geq 0} \mathbb{E}_{t \wedge \xi}[F(Z_s)] < \infty$, hence $(Z_{t \wedge \xi})_{t \geq 0}$ is uniformly integrable.

To see that $\lim_{t \to \infty} \zeta_t = \infty$, note that $\mathbb{P}\left(\int_0^\infty \mathbb{1}_{[-1,1]}(Z_t) \, dt = \infty \right) = 1$ and that F'' is bounded away from 0 on $[-1, 1]$.

Finally it remains to show that X is uniformly integrable. To see this we apply again the de la Vallée-Poussin theorem to obtain an increasing, super-linear, convex function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\int g \circ F \, d\mu < \infty$. Since $g \circ F$ is convex we find that $\sup_t \mathbb{E}_{t \wedge \xi}[g \circ F(Z_s)] < \infty$. Thus $(F(Z_{\xi \wedge t}))_{t \geq 0}$ is uniformly integrable and this carries over to X. $\square$

Definition 8.4. *Let $X : C(\mathbb{R}_+) \to \mathbb{R}$ be a measurable function which is bounded or positive. Then we define $\mathbb{E}[X|\mathcal{F}_t^0]$ to be the unique $\mathcal{F}_t^0$-measurable function satisfying*

$$\mathbb{E}[X|\mathcal{F}_t^0](\omega) = \int X((\omega_{\restriction[0,t]} \otimes \omega') \, d\mathbb{P}_{\omega(t)}(\omega').$$

Then the natural analogue of Proposition 4.11 holds by the Feller property of X:

Proposition 8.5. *Let $X \in C_b(C(\mathbb{R}_+))$. Then $X_t(\omega) := \mathbb{E}[X|\mathcal{F}_t^0](\omega)$ defines an S-continuous martingale. We denote this martingale by X^M.*

Proof. By the Feller property, we have for any continuous and bounded function X and any sequence $x_n \to x$ that also $\int X \, d\mathbb{P}_{x_n} \to \int X \, d\mathbb{P}_x$. Together with the argument used in the derivation of Proposition 4.11 this yields the desired result. $\square$

This allows us to prove the following duality result which we state in the case $\lambda = \delta_0$ for ease of exposition. The case of a general starting law follows as in Section 5.6.

Theorem 8.6. *Let $\gamma : S \to \mathbb{R}$ be S-lower semi-continuous and bounded from below. Suppose that Assumption 8.2 holds. Put*

$$P_\gamma(\mathbb{P}, \mu) := \inf_{\xi \in \mathsf{PRST}(\mu)} \int \gamma \, d\xi = \inf_{\xi \in \mathsf{RST}(\mu)} \int \gamma \, d\xi.$$

Let $\mathsf{DC}(\gamma)$ be the set of all pairs (ψ, φ) such that $\gamma(\omega, t) \geq \varphi(\omega, t) + \psi(\omega(t))$, where $\psi \in C(\mathbb{R})$ and φ is a $\mathbb{P}$-semimartingale with decomposition $\varphi = M^\varphi + A^\varphi$ where M^φ is an S-continuous and bounded $\mathbb{P}$-martingale starting at zero and A^φ is a decreasing process satisfying $\inf_{\xi \in \mathsf{RST}(\mu)} \int A^\varphi \, d\xi \geq 0$. Put

$$D_\gamma(\mathbb{P}, \mu) := \sup_{(\psi, \varphi) \in \mathsf{DC}(\gamma)} \int \psi \, d\mu.$$

Then, it holds that $P_\gamma(\mathbb{P}, \mu) = D_\gamma(\mathbb{P}, \mu)$. Moreover, in case (1) of Assumption 8.2, the process A^φ may be assumed to be zero at the expense of assuming that $M^\varphi_{t \wedge \xi}$ is only uniformly integrable for all $\xi \in \mathsf{RST}(\mu)$.

Proof. Consider first case (2) of Assumption 8.2. Let $G(t)$ be an increasing, diverging function such that $\sup_{\xi \in \mathsf{PRST}(\mu)} \mathbb{E}_\xi[G(T)] =: V_G < \infty$, and note that the set

$$\mathsf{TM}^V(\mathbb{P}, \mu) := \{\pi \in \mathcal{P}(C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}) : \mathsf{proj}_{C(\mathbb{R}_+)}(\pi) = \mathbb{P}, \mathsf{proj}_{\mathbb{R}}(\pi) = \mu, \mathbb{E}_\pi[G(T)] \leq V_G\}$$

is compact. This allows us to establish the non-adapted duality result. Putting, with c as in (5.6),

$$\mathsf{DC}^V(c) := \left\{ (\psi, \varphi) : \begin{array}{l} \varphi \text{ is an } S\text{-continuous bounded } \mathbb{P}\text{-martingale}, \ \psi \in C_b(\mathbb{R}), \ \exists \alpha \geq 0, \\ \varphi_t(\omega) + \psi(y) - \alpha(G(t) - V_G) \leq c(\omega, t, y), \text{ for } \omega \in \Omega, y \in \mathbb{R}, t \in \mathbb{R}_+ \end{array} \right\}$$

we can derive the corresponding version of Proposition 5.8. Finally, we have to note that $-\alpha(G(t) - V_G)$ is a decreasing process as claimed to deduce that

$$\mathcal{D}_c^V(\mathbb{P}, \mu) \leq D_\gamma(\mathbb{P}, \mu)$$

proving the claim.

To show case (1) of Assumption 8.2, we argue along the lines above, noting that we can replace the condition $\mathbb{E}_\pi[G(T)] \leq V_G$ by $\mathbb{E}[\zeta_\pi] \leq V$, so

$$\mathsf{TM}^V(\mathbb{P}, \mu) := \{ \pi \in \mathcal{P}(C(\mathbb{R}_+) \times \mathbb{R}_+ \times \mathbb{R}) : \mathsf{proj}_{C(\mathbb{R}_+)}(\pi) = \mathbb{P}, \mathsf{proj}_{\mathbb{R}}(\pi) = \mu, \mathbb{E}[\zeta_\pi] \leq V \}.$$

Observe that this set is compact: Suppose $\varepsilon > 0$. Since $\zeta_t \to \infty$, $\mathbb{P}$-a.s., we deduce the existence of a $K > 0$ such that $\mathbb{P}(\zeta_K < 2V/\varepsilon) < \varepsilon/2$. Hence $\mathbb{E}[\zeta_\pi] \leq V$ implies $\pi(T \geq K) < \varepsilon$, and $\mathsf{TM}^V(\mathbb{P}, \mu)$ is indeed compact.

Finally, observe that we can use the S-continuous martingale $h(\omega(t)) - \zeta_t$ to replace $\omega(t)^2 - t$ in (5.9) to deduce that

$$\mathcal{D}_c^V(\mathbb{P}, \mu) \leq D_\gamma(\mathbb{P}, \mu). \qquad \square$$

Apart from the abstract theory the ingredients to prove Theorem 6.5 are Proposition 6.6 and Proposition 6.7. The latter builds on Proposition 4.11 to identify certain hitting times as $\mathcal{F}^+$-stopping times. We just proved the analogue of Proposition 4.11 (Proposition 8.5) and in fact Propositions 6.6, 6.7 carry over to the present setup. We thus obtain:

Theorem 8.7. *Assume that $\gamma : S \to \mathbb{R}$ is Borel-measurable, the optimization problem (5.1) is well-posed and that $v \in \mathsf{RST}(\mu)$ is an optimizer. Then there exists a γ-monotone Borel set $\Gamma \subseteq S$ which supports v in the sense that $r(v)(\Gamma) = 1$.*

8.1. **Examples: One-dimensional Diffusions.** Let Z_t be a regular (time-homogenous) one-dimensional diffusion on an interval $I \subseteq \mathbb{R}$, with inaccessible or absorbing endpoints (see [45] for the relevant definitions and terminology) and $Z_0 \sim \lambda$, some $\lambda \in \mathcal{P}(I)$. In particular, Z_t is a Feller process ([45, Proposition V.50.1]). Then (on a possibly enlarged probability space) there exists a scale function $s(x)$ and a continuous, strictly increasing time change A_t such that $B_t = s(Z_{A_t})$ is a Brownian motion up to the exit of $s(I^\circ)$. Recalling the discussion in [11, Section 5], with the obvious extension of our notation, it is clear that there exists a stopping time $\xi \in \mathsf{RST}(\mu; Z)$ if and only if there exists a stopping time $\xi' \in \mathsf{RST}(s(\mu); B)$ such that $\xi'(\llbracket 0, \tau_{s(I)} \rrbracket) = 1$, where $\tau_{s(I)} = \inf\{t \geq 0 : B_t \notin s(I^\circ)\}$. Write A_t^{-1} to be the inverse of A_t, so $A_{A_t}^{-1} = t$. We now consider three cases:

- Suppose $s(I^\circ) = (a, b)$ for $a, b \in \mathbb{R}$. Then it follows from [10, Theorems 17 and 22] that $\mathsf{RST}(\mu; Z)$ is non-empty if and only if $s(\lambda)$ precedes $s(\mu)$ in convex order, and in fact, any $\xi \in \mathsf{RST}(Z)$ with $Z_\xi \sim \mu$ is minimal (so $\xi \in \mathsf{RST}(\mu, Z)$).
- Suppose $s(I^\circ) = (a, \infty)$ for $a \in \mathbb{R}$, and $s(\lambda), s(\mu)$ are integrable measures. Write $m_\lambda = \int s(y) \lambda(dy)$, and $m_\mu = \int s(y) \mu(dy)$. Then it follows from Theorems 17 and 22 and the discussion at the top of p. 245 of [10] that $\mathsf{RST}(\mu; Z)$ is non-empty if and only if $-\int |s(y) - x| \mu(dy) \leq -\int |s(y) - x| \lambda(dy) + (m_\lambda - m_\mu)$ for all $x \geq a$, or equivalently, that $\int (s(y) - x)_+ \mu(dy) \leq \int (s(y) - x)_+ \lambda(dy)$ for all $x \geq a$. Again, any $\xi \in \mathsf{RST}(Z)$ with $Z_\xi \sim \mu$ is minimal. By symmetry, similar results for the case where $s(I^\circ) = (-\infty, b)$ for $b \in \mathbb{R}$ can be given.
- Suppose $s(I^\circ) = (-\infty, \infty)$, and $\int (s(y))^2 \lambda(dy), \int (s(y))^2 \mu(dy) < \infty$. Then we are in the classical case, and a stopping time $\xi \in \mathsf{RST}(Z)$ is minimal if and only if $Z_\xi \sim \mu$, and $\mathbb{E}[A_\xi^{-1}] < \infty$. In particular, compactness of $\mathsf{RST}(\mu; Z)$ follows directly from compactness of $\mathsf{RST}(s(\mu); B)$. Further, if the scale function s is suitably differentiable, one can show that $X_t = s(Z_t)^2 - A_t^{-1}$ is an S-continuous martingale, and under the condition that $\int (s(y))^2 \lambda(dy), \int (s(y))^2 \mu(dy) < \infty$, we

deduce that $X_{t \wedge \xi}$ is uniformly integrable for all $\xi \in \mathsf{RST}(\mu; Z)$. In particular, (1) of Assumption 8.2 is satisfied.

More generally, when only the integrals $\int s(y) \lambda(dy)$ and/or $\int s(y) \mu(dy)$ are finite, we are in the situation of Proposition 8.3 and Assumption 8.2 (1) is satisfied.

Remark 8.8. Observe that none of the constructions described in Section 7.2 rely on fine properties of Brownian motion — the main properties used are the continuity of paths, the Strong Markov property, and the regularity and diffusive nature of paths (that the process started at x immediately returns to x, and immediately enters the sets (x, ∞) and $(-\infty, x)$. It follows that all the given constructions extend to the case of regular one-dimensional diffusions described above.

8.1.1. *Brownian motion with drift.* Let $Z_t = B_t + at$ for some $a < 0$ with $Z_0 \sim \lambda$, and $I = (-\infty, \infty)$. Then a possible choice of the scale function is $s(x) = \exp(-2ax)$, and $s(I^\circ) = (0, \infty)$. Let $\lambda, \mu \in \mathbb{P}(\mathbb{R})$ be such that $s(\lambda), s(\mu)$ are integrable, and suppose

$$\int (\exp(-2ay) - x)_+ \, \mu(dy) \le \int (\exp(-2ay) - x)_+ \, \lambda(dy),$$

for all $x \ge 0$. By the arguments above, there exists an embedding and all stopping times embedding μ are minimal. Then the set $\mathsf{RST}(\mu; Z)$ is compact as can be seen by the following estimate inserted in the proof of Theorem 4.28. Fix $\varepsilon > 0$ and take $K > 0$ such that $\mu((-\infty, -K)) \le \varepsilon/4$. Then there is $R > 0$ such that

$$\mathbb{P}_\lambda(\exists \, R' \ge R : Z_{R'} \ge -K) \le \varepsilon/4.$$

Then $\xi \in \mathsf{RST}(\mu)$ implies that $\xi(T > R) \le \varepsilon/2$.

8.1.2. *Geometric Brownian motion.* Let Z be a geometric Brownian motion and μ be concentrated on the *positive reals* $(0, \infty)$. Then the compactness of $\mathsf{RST}(\mu; Z)$ follows from the compactness in the case of Brownian motion with drift as $\exp : \mathbb{R} \to (0, \infty)$ is a homeomorphism. Similarly, conditions for the existence and minimality of $\xi \in \mathsf{RST}(\mu; Z)$ follow directly from the case of Brownian motion with drift, or more generally, from the observation that Z is a regular diffusion.

8.1.3. *Three-dimensional Bessel process.* Let $Z = \|B\|$ for a three-dimensional Brownian motion $(B_t)_{t \ge 0}$ (or d-dimensional with $d \ge 3$) with $Z_0 \sim \lambda$. Let $\mu \in \mathcal{P}((0, \infty))$ be such that there exists at least one embedding. Then any embedding is minimal and $\mathsf{RST}(\mu; Z)$ is compact. This can be seen by similar argument to the case of Brownian motion with drift, since B_t is transient in dimension three and higher. Indeed, fix $\varepsilon > 0$ and take $K > 0$ such that $\mu((K, \infty)) \le \varepsilon/4$. By the transience of B_t there is $R > 0$ such that

$$\mathbb{P}(\exists \, R' \ge R : Z_{R'} \le K) \le \varepsilon/4,$$

which implies that $\xi(T > R) \le \varepsilon/2$ implying the compactness of $\mathsf{RST}(\mu; Z)$ by a straightforward modification of Theorem 4.28.

8.1.4. *Ornstein-Uhlenbeck processes.* Let Z be an Ornstein-Uhlenbeck process, given for example as the solution to the SDE $dZ_t = -Z_t \, dt + dW_t$. Then Z_t is a regular diffusion on $I = (-\infty, \infty)$ with scale function given (up to constants) by $s'(x) = \exp(x^2)$. Then $s(I^\circ) = (-\infty, \infty)$. Suppose λ, μ are measures on $\mathbb{R}$ with $s(\lambda), s(\mu)$ square integrable, and in convex order. Then $\mathsf{RST}(\mu; Z)$ is compact and $\xi \in \mathsf{RST}(\mu; Z)$ if and only if $Z_\xi \sim \mu$ and $\mathbb{E}[A_\xi^{-1}] < \infty$.

8.1.5. *The Hoeffding-Frechet coupling as a very particular Root solution.* Let Z be the deterministic process given by $dZ_t = dt$ started in $Z_0 \sim \lambda$. Z is not a regular diffusion, however Assumption 8.2 (2) is easily checked. Let μ be another probability and assume for simplicity that $\max \operatorname{supp} \lambda \le \min \operatorname{supp} \mu$. Then the Root solution minimizes $\mathbb{E}[\tau^2]$. But note also that since $\tau = Z_\tau - Z_0$, this minimization problem corresponds precisely to finding the joint distribution (Z_0, Z_τ) which minimizes $\mathbb{E}[(Z_\tau - Z_0)^2]$: the classical transport problem in the most simple setup. Specifically, the Root solution for the particular case of

the process Z corresponds precisely to the monotone (Hoeffding-Frechet) coupling. In the same fashion the Rost solution corresponds to the co-monotone coupling between λ and μ.

References

[1] B. Acciaio, M. Beiglböck, F. Penkner, W. Schachermayer, and J. Temme. A trajectorial interpretation of doob's martingale inequalities. *The Annals of Applied Probability*, 23(4):1494–1505, 2013.

[2] D. R. Adams and L. I. Hedberg. *Function spaces and potential theory*, volume 314 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1996.

[3] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In *Optimal transportation and applications (Martina Franca, 2001)*, volume 1813 of *Lecture Notes in Math.*, pages 123–160. Springer, Berlin, 2003.

[4] J. Azéma and M. Yor. Une solution simple au problème de Skorokhod. In *Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg, 1977/78)*, volume 721 of *Lecture Notes in Math.*, pages 90–115. Springer, Berlin, 1979.

[5] J. R. Baxter and R. V. Chacon. Compactness of stopping times. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40(3):169–181, 1977.

[6] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer. Optimal and better transport plans. *J. Funct. Anal.*, 256(6):1907–1927, 2009.

[7] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: A mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.

[8] S. Bianchini and L. Caravenna. On optimality of c-cyclically monotone transference plans. *C. R. Math. Acad. Sci. Paris*, 348(11-12):613–618, 2010.

[9] B. Bouchard and M. Nutz. Arbitrage and Duality in Nondominated Discrete-Time Models. *Ann. Appl. Probab., to appear*, 2013.

[10] A. M. G. Cox. Extending Chacon-Walsh: minimality and generalised starting distributions. In *Séminaire de probabilités XLI*, volume 1934 of *Lecture Notes in Math.*, pages 233–264. Springer, Berlin, 2008.

[11] A. M. G. Cox and D. Hobson. Skorokhod embeddings, minimality and non-centred target distributions. *Probability Theory and Related Fields*, 135:395–414, 2005.

[12] A. M. G. Cox and J. Obłój. Robust hedging of double touch barrier options. *SIAM J. Financial Math.*, 2:141–182, 2011.

[13] A. M. G. Cox and G. Peskir. Embedding Laws in Diffusions by Functions of Time. *Ann. Probab., to appear*, 2014.

[14] A. M. G. Cox and J. Wang. Root's Barrier: Construction, Optimality and Applications to Variance Options. *Ann. Appl. Probab.*, 23(3):859–894, 2013.

[15] C. Dellacherie and P.-A. Meyer. *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1978.

[16] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields*, 160(1-2):391–427, 2014.

[17] Y. Dolinsky and H. M. Soner. Martingale optimal transport in the Skorokhod space. *ArXiv e-prints*, April 2014.

[18] Y. Dolinsky and M. H. Soner. Robust hedging with proportional transaction costs. *Finance Stoch.*, 18(2):327–347, 2014.

[19] Neil Falkner. The distribution of Brownian motion in $\mathbf{R}^n$ at a natural stopping time. *Adv. in Math.*, 40(2):97–127, 1981.

[20] A. Galichon, P. Henry-Labordère, and N. Touzi. A Stochastic Control Approach to No-Arbitrage Bounds Given Marginals, with an Application to Lookback Options. *Ann. Appl. Probab.*, 24(1):312–336, 2014.

[21] W. Gangbo and R. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[22] P. Gassiat, A. Mijatovic, and H. Oberhauser. An integral equation for Root's barrier and the generation of Brownian increments. *Ann. Appl. Probab., to appear*, 2014.

[23] Oblój J. Spoida P. Henry-Labordère, P. and N. Touzi. Maximum Maximum of Martingales Given Marginals. *Ann. Appl. Probab., to appear*, 2012.

[24] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and associated martingales, with explicit constructions*, volume 3 of *Bocconi & Springer Series*. Springer, Milan; Bocconi University Press, Milan, 2011.

[25] D. Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2:329–347, 1998.

[26] D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 267–318. Springer, Berlin, 2011.

[27] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Mathematical Finance*, 22(1):31–56, December 2012.

[28] S. Jacka. Doob's inequalities revisited: A maximal h^1-embedding. *Stochastic processes and their applications*, 29(2):281–290, 1988.

[29] H. G. Kellerer. Duality theorems for marginal problems. *Z. Wahrsch. Verw. Gebiete*, 67(4):399–432, 1984.

[30] J. Kiefer. Skorohod embedding of multivariate RV's, and the sample DF. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 24(1):1–35, 1972.

[31] M. Knott and C. Smith. On Hoeffding-Fréchet bounds and cyclic monotone relations. *J. Multivariate Anal.*, 40(2):328–334, 1992.

[32] M. Knott and C. S. Smith. On the optimal mapping of distributions. *J. Optim. Theory Appl.*, 43(1):39–49, 1984.

[33] G. Last, P. Mörters, and H. Thorisson. Unbiased shifts of Brownian motion. *Ann. Probab.*, 42(2):431–463, 2014.

[34] R. M. Loynes. Stopping times on Brownian motion: Some properties of Root's construction. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 16:211–218, 1970.

[35] P.-A. Meyer. Convergence faible et compacité des temps d'arrêt, d'après baxter et chacon. In *Séminaire de probabilités*, volume XII of *Lecture Notes in Mathematics, 649*, pages 411–423. Springer Berlin Heidelberg, 1978.

[36] I. Monroe. On embedding right continuous martingales in Brownian motion. *Ann. Math. Statist.*, 43:1293–1311, 1972.

[37] A. Nikeghbali and M. Yor. A definition and some characteristic properties of pseudo-stopping times. *The Annals of Probability*, 33(5):1804–1824, September 2005.

[38] H. Oberhauser and G. dos Reis. Root's barrier, viscosity solutions of obstacle problems and reflected FBSDEs. *ArXiv e-prints*, January 2013.

[39] J. Obłój. The Skorokhod embedding problem and its offspring. *Probab. Surv.*, 1:321–390, 2004.

[40] J. Obłój, P. Spoida, and N. Touzi. Martingale Inequalities for the Maximum via Pathwise Arguments. *ArXiv e-prints*, September 2014.

[41] E. Perkins. The Cereteli-Davis solution to the H^1-embedding problem and an optimal embedding in Brownian motion. In *Seminar on stochastic processes, 1985*, pages 172–223. Springer, 1986.

[42] A. Pratelli. On the sufficiency of *c*-cyclical monotonicity for optimality of transport plans. *Math. Z.*, 258(3):677–690, 2008.

[43] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[44] R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math.*, 17:497–510, 1966.

[45] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes and martingales: Volume 2, Itô calculus*, volume 2. Cambridge university press, 2000.

[46] D. H. Root. The existence of certain stopping times on Brownian motion. *Ann. Math. Statist.*, 40:715–718, 1969.

[47] H. Rost. The stopping distributions of a Markov Process. *Invent. Math.*, 14:1–16, 1971.

[48] H. Rost. Skorokhod stopping times of minimal variance. In *Séminaire de Probabilités, X (Première partie, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)*, pages 194–208. Lecture Notes in Math., Vol. 511. Springer, Berlin, 1976.

[49] L. Rüschendorf. Fréchet-bounds and their applications. In *Advances in probability distributions with given marginals (Rome, 1990)*, volume 67 of *Math. Appl.*, pages 151–187. Kluwer Acad. Publ., Dordrecht, 1991.

[50] L. Rüschendorf. Optimal solutions of multivariate coupling problems. *Appl. Math. (Warsaw)*, 23(3):325–338, 1995.

[51] L. Rüschendorf. On *c*-optimal random variables. *Statist. Probab. Lett.*, 27(3):267–270, 1996.

[52] W. Schachermayer and J. Teichmann. Characterization of optimal transport plans for the Monge-Kantorovich problem. *Proc. Amer. Math. Soc.*, 137(2):519–529, 2009.

[53] A. V. Skorohod. *Issledovaniya po teorii sluchainykh protsessov (Stokhasticheskie differentsialnye uravneniya i predelnye teoremy dlya protsessov Markova)*. Izdat. Kiev. Univ., Kiev, 1961.

[54] A. V. Skorokhod. *Studies in the theory of random processes*. Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.

[55] H. Strasser. *Mathematical theory of statistics*, volume 7 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1985. Statistical experiments and asymptotic decision theory.

[56] P. Vallois. Le probleme de Skorokhod sur ℝ: une approche avec le temps local. In *Séminaire de Probabilités XVII 1981/82*, pages 227–239. Springer, 1983.

[57] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[58] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.